
The Brunswik Society

Newsletter

Vol. 31, Nov. 2016

To get information about future Newsletters and the Brunswik Society you are welcome to join (free of charge) the email list of the Brunswik Society at:

<http://www.brunswik.org/>

ISSN 2296-9926

Table of Contents

Foreword

Kaufmann, E., Athanasou, J. A., & Hamm, R. M...... 4

Contributions

Beckstead, J. W.
Introducing the Bifocal Lens Model and Equation..... 5

Dhami, M. K.
Using Brunswikian Methods to Get Inside the Judge’s Mind..... 7

Giraudeau, C., Musielak, C., & Chasseigne, G.
Inhibition among Older Adults: A Brunswikian View..... 8

González-Vallejo, C.
Brunswik Lens Model Analyses to Understanding Judgments of Nutrition..... 11

Hamm, R. M., & Yang, H.
Comparison of Logistic and Linear Regression for Lens Models Predicting Dichotomous Judgments..... 12

Hoffrage, U., & Marewski, J. N.
Modeling and Aiding Intuition: In the Twilight between the Conscious and the Unconscious 14

Holzworth, J. R., Stewart, T. R., & Mumpower, J. L.
Learning to Make Selection and Detection Decisions: The Role of Task Uncertainty..... 18

Kaufmann, E., & Wittmann, W. W.
A Hammond-Brunswik Orientated Evaluation: The Success of Linear Bootstrapping Models as an Example..... 19

Kausel, E. E.
Using Brunswik’s Lens Model to Nail the (Unstructured) Interview..... 21

Keuschnigg, M., & Ganser, C.
Brunswik’s Lens Model Approach in Agent-based Simulations on Crowd Wisdom..... 23

Koch, A., & Imhoff, R.
Towards an Ecologically Valid Model of Spontaneous Stereotypes about Groups..... 25

Kunc, M. H.
Modeling Managerial Decision Making: A Brunswikian Perspective..... 28

Leuker, C., Pleskac, T. J., Pachur, T., & Hertwig, R.
How the Risk-Reward Relationship Shapes Decisions Under Risk and Uncertainty..... 30

Nadler, I.
Clinical Judgments in Acute Medical Care..... 32

Nuamah, J. K., & Seong, Y.
A Neuroergonomics Approach to Cognition in Judgment and Decision Making..... 34

Pak, R., Rovira, E., & McLaughlin, A. C.
The Importance of Representative Design in Human Factors Automation Research..... 37

Powell, D.
Which Anxiety Cues do Employment Interviewers Detect?..... 39

Rogers, T., ten Brinke, L., & Carney, D. R.
Unacquainted Callers can Predict which Citizens will Vote Over and Above Citizens'
Stated Self-Predictions..... 41

Sjödahl, L.
What is Represented in a Vignette?..... 44

Tomassetti, A. J., Dalal, R. S., & Kaplan, S. A.
Policy Capturing is More Resistant to Faking than Other More Traditional Self-Report
Techniques..... 48

Imprint

The Brunswik Society Newsletter, ISSN 2296-9926
is published yearly
Editorial staff:
Kaufmann, E., Athanasou, J. A., & Hamm, R. M.

Contact:

The Brunswik Society Newsletter
c/o University of Zurich
Institute of Education, University of Zurich,
Kantonschulstr. 3
CH-8006 Zurich
Switzerland

Foreword

It is a great pleasure to present the newest Brunswik Society Newsletter covering recent research which is inspired by the work of Egon Brunswik or Ken Hammond.

Although Brunswik's last publication appeared in 1966, his thoughts still inspire young researchers (see the contributions by Leuker, Koch, Nadler). They challenge university teaching (see the contribution by Kausel).

Moreover, in this issue you will find several contributions applying the lens model to different fields (see the work of González-Vallejo or Giraudeau or Holzworth). It is also applied to new fields such as neuroergonomics (see Nuamah). Within the Newsletter there are also contributions that take methodological aspects of the lens model into consideration and suggest further refinements (see Beckstead's and Hamm's contributions).

Beside Brunswik's inspiration to Hammond for the field of judgment and decision-making, Brunswik was also well-known for his representative design approach, and several contributions take this up in different fields (see e.g., Pak, Koch, Sjö Dahl).

Taken together, our recent Newsletter clearly shows the continued relevance of Brunswikian research – within the field of judgment and decision-making, but also the power of this approach to the methodological challenges faced today within science. Our newsletter also highlights the activity of the community to improve the original work and to follow the original spirit of Brunswik/Hammond.

To keep-up-to-date with thoughts and discussions for further research, please follow the Brunswik Society mailing list, which you will find at: www.brunswik.org. We are looking forward to hearing your thoughts, and hope that gathering these ideas here stimulates discussion, disagreement, and collaboration among researchers.

We hope that the richness of the contributions included in this year's Newsletter inspires your research.

Many thanks to all authors for their contributions!

Sincerely,
Esther Kaufmann, James A. Athanasou and Robert M. Hamm

We greatly thank Martin Kunc for his work as a reviewer. If you're interested to support the editorial team of the Brunswik Society Newsletter and to be involved in the next Brunswik Society Newsletter let us know by email (esther.kaufmann@gmx.ch).

Thank you to Tom Stewart, the webmaster of the Brunswik Society, for providing web access to the Newsletter.

Contributions

Introducing the Bifocal Lens Model and Equation

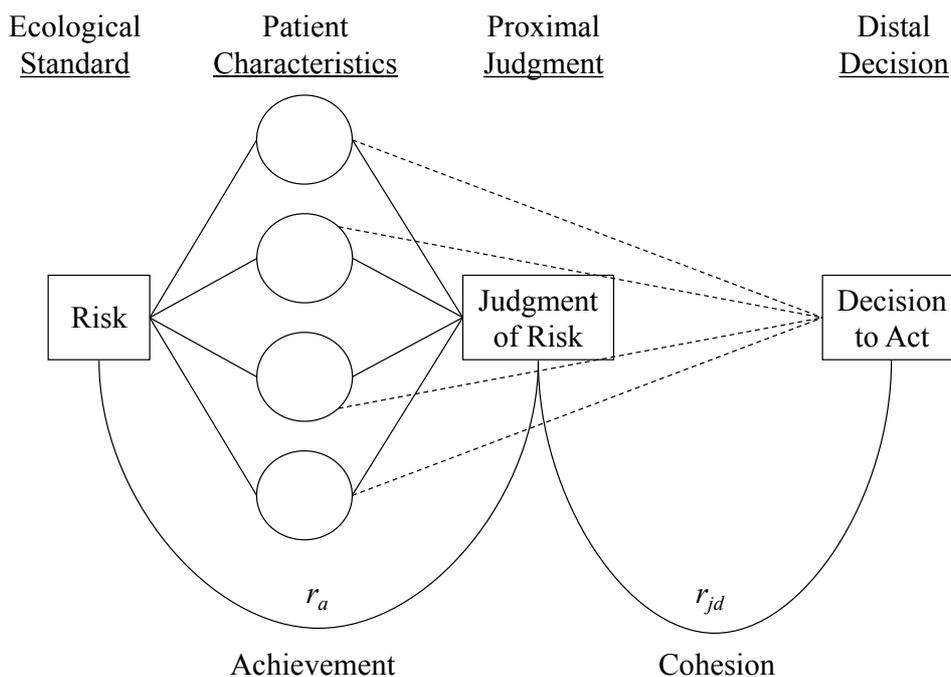
Jason W. Beckstead
University of South Florida, US

Contact: jbeckste@health.usf.edu

Brunswick's Lens Model and lens model equation (*LME*) have been applied extensively in medical decision-making research. In many patient encounters, clinicians face the dual challenge of formulating a judgment of patient risk for some adverse outcome and making a yes/no decision regarding a particular risk-reducing treatment option. For example, judging a hospital patient's risk for acute deterioration and decision to alert the hospital's rapid response team. When asked to review the same set of patient profiles, clinicians can show variation in the degree to which their quantitative judgments of patient risk are correlated with their yes/no decisions to take a specific risk-reducing action. I introduce the term, *cohesion*, to describe how well these two related cognitive responses (quantitative judgments and dichotomous decisions) correlate, or hang together within an individual. Cohesion implies that for a given clinician, higher likelihoods of taking risk-reducing actions will co-occur with higher judgments of patient risk, while lower likelihoods of taking risk-reducing actions will co-occur with lower judgments of risk.

This year I have published an article developing and applying a novel form of Brunswick's Lens Model and *LME* to decompose cohesion (Beckstead, 2016). In the paper I demonstrate how this approach can be used to gain insight into individual differences in the linkage between clinical risk judgments and treatment-related decisions. The model is "bifocal" in that it focuses on two sets of linked responses (quantitative judgments of risk and dichotomous decisions to take a risk-reducing action) from the same individual. This type of dual-response task can be represented with the Bifocal Lens Model shown in the Figure below.

Consistent with the classic Lens Model, the task ecology is represented on the left-hand side and the subjective responses from the clinician are on the right-hand side. Towards the center of the figure are the cues, in this case a set of patient characteristics. The lines originating from the cues converge to the ecological criterion (Risk) and to the subjective judgment of risk, in much the same way that light passing through a lens can be focused at a point lying on either side. Achievement (r_a), or the accuracy of the individual's judgments compared with the ecological criterion, is represented by the arc in the lower left-hand side of the figure and is the correlation coefficient decomposed in the classic *LME*.



In the Bifocal Lens Model the metaphor is extended to encompass two subjective focal points on the right-hand side, one proximal and one distal. As the figure illustrates, both the proximal judgment (j) of risk and the distal decision (d) to act are modeled on the same set of cues. Cohesion (r_{jd}), or the correlation between the two sets of responses from an individual, is represented by the arc in the lower right-hand side of the figure.

The bifocal lens model equation (*BiLME*) links the results from two regression models, one linear and one logistic, based on the same set of cases examined by the same individual. In the *BiLME* the degree of cohesion (r_{jd}) between a clinician's judgments of patient risk and decisions to take a risk-reducing action is modeled as a function of internal policy matching, G_{jd} , or the extent to which he/she used the same cue weighting strategy when formulating these two responses, R_j , the amount of cognitive control in his/her judgment policy, R_d , the amount of cognitive control in his/her decision policy, and the correlation between the residuals from the two regression equations, C_1 . The *BiLME* is developed fully in the paper and data from two judgment analysis studies are analyzed to illustrate how individual differences in cohesion can be explained by individual differences in the *BiLME* parameters.

Across clinicians cohesion varied due to individual differences in internal policy matching (G_{jd}), individual differences in cognitive control (R_j and R_d), and individual differences in reliance on a subjective threshold which was independent of cue utilization (represented by C_1). In contrast, individual differences in cohesion showed little to no relationship with individual differences in achievement. What makes this finding particularly interesting is that it suggests some individuals can show strong connections between their decisions and their judgments regardless of whether their judgments are accurate. And conversely, individuals who can accurately judge risk may not always rely on their judgments when making risk-reducing decisions.

Reference:

Beckstead, J. W. (2016). The bifocal lens model and equation: Examining the linkage between clinical judgments and decisions. *Medical Decision Making*. doi:10.1177/0272989X16674196

**Using Brunswikian Methods
to Get Inside the Judge's Mind**

Mandeep K. Dhami
Middlesex University, UK

Contact: m.dhami@mdx.ac.uk

My PhD student, Ian Belton, and I have been working on a paper that critiques past research on judicial decision-making. We point out that the empirical evidence is often dismissed as irrelevant, and the judiciary, legal policy-makers and the public remain largely unconvinced that the status quo needs much improving. These audiences claim: (1) the scientific findings lack validity since researchers did not study judges making decisions on real cases, and (2) researchers have not pinpointed the psychological processes of any specific judge because they analyzed data over judges and/or used statistical models lacking in psychological plausibility.

We review these two grounds for appeal against the scientific research on judicial decision-making, and note that it appears researchers' choices of data collection methods and analytic techniques may, indeed, be inappropriate for understanding the phenomena. We offer two remedies from the sphere of (neo-) Brunswikian decision-making research: collecting data on judicial decision-making using representative experimental design, and analyzing individual judges decision data using more psychologically plausible models (e.g., fast and frugal heuristics rather than regression models). Used together, we believe these solutions can help researchers better understand and improve legal decision-making.

Therefore, Ian and I are directly appealing to researchers in the fields of Psychology, Law and Criminology to change their research habits. We will see what transpires...

Moreover, my recent publication maybe also of interest for society members:

Dhami, M. K., Belton, I., & Goodman-Delahunty, J. (2015). Quasi-rational models of sentencing. *Journal of Applied Research on Memory and Cognition*, 4, 239-247.

Cognitive continuum theory points to the middle-ground between the intuitive and analytic modes of cognition, called quasirationality. In the context of sentencing, we discuss how legal models prescribe the use of different modes of cognition. These models aim to help judges perform the cognitive balancing act required between factors indicating a more or less severe penalty for an offender. We compare sentencing in three common law jurisdictions (i.e., Australia, the US, and England and Wales). Each places a different emphasis on the use of intuition and analysis; but all are quasirational. We conclude that the most appropriate mode of cognition will likely be that which corresponds best with properties of the sentencing task. Finally, we discuss the implications of this cognition-task correspondence approach for researchers and legal policy-makers.

**Inhibition among Older Adults:
A Brunswikian View**

Caroline Giraudeau, Céline Musielak & Gérard Chasseigne
François-Rabelais University,
France

Contact: gerard.chasseigne@univ-tours.fr

Introduction

The aim of the present study (Giraudeau et al., 2016) was to explore the ability of older people to learn the strength of direct linear relationships when only certain cues are valid for predicting a criterion. To this end, we used the MCPL paradigm (Holzworth, 2001). Inhibitory theory (Hasher & Zacks, 1988) proposes that older people are less efficient than younger adults in inhibiting irrelevant information. Several studies show that the age-related differences in the inhibition performance increase with the cognitive demands (Aguirre et al., 2014; Zacks et al., 1996), although some of them show only a partial reduction in inhibitory capacities (Andrés et al., 2004).

Voluntary suppression of information in memory is usually explored using paradigms in which participants are explicitly asked to forget or to remember previously learned items. In everyday life, people are seldom instructed to suppress information in working memory. They usually have to select items to be discounted by themselves. This process can be studied using a functional learning approach where people learn functional relations between events. The ability to detect and learn these relations clearly has a strong adaptive value (Hammond & Stewart, 2001).

Although only a few sets of studies using the functional learning paradigm have been conducted in older adults (Chasseigne et al., 1997, 1999, 2001, 2002; Musielak et al., 2006, 2014), they present a picture of older people whose learning capacities enable them to detect functional relations in the learning environment almost as well as young people, and to learn rapidly from them how multiple events are related to each other.

In the present study, the performance of adults of different ages was assessed and compared under four learning conditions differing in the number of valid cues (from all four to just one). Given the participants' assumption of equal validity of the cues and direct cue-criterion relationships (Brehmer, 1974), and given that discounting invalid cues entails voluntary inhibition of information, which involves a considerable load on working memory and requires self-initiated activities, it could be hypothesized that the greater the inhibition effort needed in the task, the greater the difference between younger and older participants. As a result, the age-related difference in learning performance would be greater (a) under the one- than two-valid-cue condition, (b) under the two- than three-valid-cue condition, and (c) under the three- than four-valid-cue condition. With four valid cues, differences should be minimal.

However, an alternative view that takes into account the context in which inhibition occurs (Rush, Barch, & Braver, 2006) could lead to a different hypothesis. In the present study, participants had to detect the set of cues that were valid in a four-cue setting, without knowing in advance the number of valid cues. Thus, when only one cue is valid, there are only four possibilities: the valid cue is either Cue A, Cue B, Cue C or Cue D. Likewise, when three cues are valid, there are also four (complementary) possibilities: the valid set of cues is composed of either Cues A, B and C, Cues A, B and D, Cues A, C and D, or Cues B, C and D. In terms of level of uncertainty, these two conditions are equivalent. By contrast, when two cues are valid, there are six possibilities (Cues AB, AC, AD, BC, BD, or CD). In addition, if the one-valid-cue condition and the three-valid-cue condition are equivalent in terms of uncertainty in the task, they are far from being equivalent in terms of working memory load. Learning which of four sets of three cues predicts a criterion places a greater burden on working memory than looking for which one of four cues predicts the criterion. According to this view, the age-related difference in learning performance would be greater (a) under the two-valid-cue condition than under the three-valid-cue condition, (b) under the three-valid-cue condition than under the one-valid-cue condition, and (c) under the one-valid-cue condition than under the all-valid-cue condition. This alternative was tested in the present study and is discussed below.

Methods

A total of 240 adults (18-90 years old) had to learn to predict the amount of drink delivered by a drink dispenser on the basis of four cues (the height of four vertical bars). The participants were randomly distributed between four experimental conditions, one valid cue, two valid cues, three valid cues and all valid cues. The measures that were calculated for each participant under each condition included the squared mean differences between judgment and criterion as an index of performance, and cue utilizations as a test of both the learning of the strength of direct linear relationships and of inhibition.

Results

The results validated the hypothesis that the level of difficulty of the four learning conditions can be better predicted from the number of possible sets of valid cues to be utilized than from the number of sets of non-pertinent cues to be inhibited. In all conditions and in each age group, cue utilizations were direct in the first block with no feedback. Older adults discounted the non-pertinent cues as well as younger adults, while participants aged over 76 only succeeded under the least demanding conditions. The presence of non-pertinent cues affected the learning of direct cues, even among the younger participants.

Discussion and conclusion

Unlike inhibition theory, this study, which was conducted in a Brunswikian framework, shows that the older adults' ability to detect (and use) valid cues in an environment containing both valid and invalid cues appeared to be fairly well preserved. The finding that the mere presence of invalid cues can affect the learning of direct cues by both older and younger adults constitutes a new result in functional learning. Learning complex relations, such as inverse relations, does not produce such results (e.g., Chasseigne et al., 1997, 2004). Future research should examine the role of invalid cues in functional learning.

References:

- Aguirre, C., Gómez-Ariza, C. J., Bajo, M. T., Andrés, P., & Mazzoni, G. (2014). Selective voluntary forgetting in young and older adults. *Psychology and Aging, 29*(1), 128–139.
- Andrés, P., Van der Linden, M., & Parmentier, F. B. R. (2004). Directed forgetting in working memory: Age-related differences. *Memory, 12*(2), 248–256.
- Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inferences tasks. *Organizational Behavior and Human Performance, 11*, 1–27.
- Chasseigne, G., Grau S., Le Gall, A., Ligneau, C., Roque, M., & Mullet, E. (2004). Aging and probabilistic learning in single- and multiple-cue tasks. *Experimental Aging Research, 30*, 23–45.
- Chasseigne, G., Grau S., Mullet, E., & Cama, V. (1999). How well do elderly people cope with uncertainty in a learning task? *Acta Psychologica, 103*, 229–238.
- Chasseigne, G., Lafon, P., & Mullet, E. (2002). Aging and rule learning: The case of the multiplicative law. *American Journal of Psychology, 115*(3), 315–330.
- Chasseigne, G., & Mullet, E. (2001). Probabilistic Functioning and cognitive aging. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 423–426). Oxford: Oxford University Press.
- Chasseigne, G., Mullet, E., & Stewart, T. (1997). Aging and multiple cue probability learning: The case of inverse relationships. *Acta Psychologica, 97*, 235–252.
- Giraudeau, C., Musielak, C., Hervé, C., Seren, D., Chasseigne, G., & Mullet, E. (2016). Aging, functional learning, and inhibition. *Experimental Aging Research, 42*(4), 329–347.
- Hammond, K. R., & Stewart, T. R. (2001). *The essential Brunswik: Beginnings, explications, applications*. Oxford: Oxford University Press.
- Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. (1975). Social Judgment Theory. In M. Kaplan & S. Schartz (Eds.), *Human judgment and decision processes*. New York: Academic Press.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 22* (pp. 193–225). New York: Academic Press.
- Holzworth, R. J. (2001). Multiple cue probability learning. In K. R. Hammond & T. R. Stewart. *The essential Brunswik: Beginnings, explications, applications* (pp. 348–349). Oxford: Oxford University Press.
- Musielak, C., Chasseigne, G., & Mullet, E. (2006). The learning of linear and nonlinear functions in younger and older adults. *Experimental Aging Research, 32*(3), 317–339.
- Musielak, C., Giraudeau, C., Chasseigne, G., & Mullet, E. (2014). To what extent does the existence of functional relations in a learning setting change the pattern of differences between younger and older adults' performances? *Experimental Aging Research, 40*(4), 455–476.
- Rush, B. K., Barch, D. M., & Braver, T. S. (2006). Accounting for cognitive aging: Context processing, inhibition or processing speed? *Aging, Neuropsychology and Cognition, 13*(3-4), 588–610.
- Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 143–156.

**Brunswick Lens Model Analyses to
Understanding Judgments of Nutrition**

Claudia González-Vallejo
Ohio University, US

Contact: gonzalez@ohio.edu

I continue to work on applying the lens model to the understanding of consumer judgment of the nutritional value of food products. This is an ever more important question in light of current FDA attempts at redefining what “healthy” foods are. In order to define the accuracy of judgments, we used a nutrition expert system algorithm known as NuVal® as the gold standard (Katz et al., 2009). Our study in the journal of *Public Health and Nutrition* (González-Vallejo & Lavins, 2015) demonstrated that US consumers’ judgments of nutrition of cereals based on NFP information was, on average, in ordinal agreement with NuVal. In addition, we showed that participants using the newly designed FDA nutrition facts panels (NFPs) did not have greater judgment accuracy, thus highlighting the importance of conducting research prior to implementing expensive policy changes.

We published our second study in the journal *Appetite* (González-Vallejo, Lavins, & Carter, 2016). Working with individual level data and using a novel methodology for extracting best models (the criticality of predictors’ technique of Azen, Budescu, & Reiser, 2001), we found low accuracy in terms of nutrition judgments of cereal and snacks. More generally, analyses of the lens model indices (e.g., judgment consistency, judgment accuracy, and knowledge of the environment) demonstrated great variability across individuals and generally low median values. Predicting lens model indices from person level characteristics (e.g., education, gender, etc.) was challenging. Significant models resulted but R-squares were generally low. Furthermore, predicting choices of products demonstrated that although participants appeared to use their judgments of nutrition to select products (in addition to the liking of the product, and the frequency of consuming it) their choices were not for foods with higher nutritional value. We concluded that the benefits of having NFP information is perhaps lower than previously assumed from studies that used self-reports or point of purchase assessments of NFP usage. Current studies are focusing on increasing accuracy as a function of front-of-package labeling, and future studies hope to contrast nutrition judgments cross-nationally (Chile and the USA).

References:

- Azen, R., Budescu, D. V., & Reiser, B. (2001). Criticality of predictors in multiple regression. *British Journal of Mathematical and Statistical Psychology*, 54(2), 201–225.
- González-Vallejo, C., Lavins, B. D., & Carter, K. A. (2016). Analysis of nutrition judgments using the Nutrition Facts Panel. *Appetite*, 105, 71–84.
- González-Vallejo, C., & Lavins, B. D. (2015). Evaluation of breakfast cereals with the current nutrition facts panel (NFP) and the Food and Drug Administration’s NFP proposal. *Public Health and Nutrition*, 19(6), 1047–1058.
- Katz, D. L., Njike, V. Y., Faridi, Z., Rhee, L. Q., Reeves, R. S., Jenkins, D. J., & Ayoob, K. T. (2009). The stratification of foods on the basis of overall nutritional quality: The overall nutritional quality index. *American Journal of Health Promotion*, 24(2), 133–143.

Comparison of Logistic and Linear Regression for Lens Models Predicting Dichotomous Judgments

Robert M. Hamm

*Department of Family and Preventive Medicine,
University of Oklahoma Health Sciences Center, Oklahoma City, OK, US*

Huiqin Yang

*Peninsula Technology Assessment Group (PenTAG),
Evidence Synthesis & Modelling for Health Improvement (ESMI),
University of Exeter Medical School, UK*

Contact: Robert-Hamm@ouhsc.edu

We have explored the use of logistic regression for creating a lens model when the predicted entity (judgment or criterion) is dichotomous (Hamm & Yang, 2016). The paper is available at <http://onlinelibrary.wiley.com/enhanced/doi/10.1002/bdm.1969>; be sure to download the ‘supporting information’ which has lots of material that the authors (at least) think is essential.

Using data from Yang’s thesis project at York studying nurse judgments about patients in the emergency department (Yang, 2009; Yang, Thompson, Hamm, Bland, & Foster, 2013), we applied several versions of the lens model equation to the dichotomous judgments (“at risk” or not) about dichotomous data (had a bad outcome, or not), comparing their performance. Linear regression applied to dichotomous judgments (and dichotomous criterion); logistic regression (conceptually more appropriate); linear regression applied to continuous “confidence adjusted” dichotomous judgments (but still used inappropriately with the dichotomous ecological criterion); and a hybrid (logistic regression applied to the dichotomous criterion, and linear regression applied to the confidence adjusted category judgments). Both the logistic regression and the confidence adjusted approaches had advantages, compared to simple linear regression.

The formula for the logistic and hybrid models is

$$r_a = G \frac{\sigma_{\tilde{y}_e} \sigma_{\tilde{y}_s}}{\sigma_{Y_e} \sigma_{Y_s}} + C_1 \frac{\sigma_{\tilde{z}_e} \sigma_{\tilde{z}_s}}{\sigma_{Y_e} \sigma_{Y_s}} + C_2 \frac{\sigma_{\tilde{y}_e} \sigma_{\tilde{z}_s}}{\sigma_{Y_e} \sigma_{Y_s}} + C_3 \frac{\sigma_{\tilde{z}_e} \sigma_{\tilde{y}_s}}{\sigma_{Y_e} \sigma_{Y_s}},$$

which is Stewart’s (2004) refinement of Cooksey’s (1996) suggested formula. Jason Beckstead in his “bifocal model” (Beckstead, 2016, and see his newsletter contribution, above) applies the same formula to data from the same judge - comparing continuous ratings and categorical judgments of the same objects.

In Yang’s data set, and in both of Beckstead’s, the advantages of the more general formulation (that can combine two logistic models, or one logistic and one linear) are rather small. But the approach is in theory more appropriate, and the procedure is feasible, so we encourage researchers to try the logistic lens model (or the hybrid with “confidence adjustment” if you have confidence ratings about each

judgment), and compare the performance of the models. In doing so, we encourage you to have your participants judge a sufficient number of cases that the models can be fit, and to make sure that the cues are not too highly intercorrelated. Only with more studies can we understand the conditions where use of the logistic lens model provides a real advantage, and the conditions where the robustness of linear regression (Dawes, 1979) lets the linear lens model be perfectly adequate.

References:

- Beckstead, J. W. (2016). The bifocal lens model and equation: Examining the linkage between clinical judgments and decisions. *Medical Decision Making*. doi:10.1177/0272989X16674196
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic Press.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571–582.
- Hamm, R. M., & Yang, H. (2016). Alternative lens model equations for dichotomous judgments about dichotomous criteria. *Journal of Behavioral Decision Making*, 1–6. doi:10.1002/bdm.1969
- Stewart, T. R. (2004). *Notes on a form of the lens model equation for logistic regression analysis*. Paper presented at the 20th Annual International Meeting of the Brunswik Society, Minneapolis, MN.
- Yang, H. (2009). *The effects of improved representative design on nurses' risk assessment judgements and confidence calibration*. (PhD-Thesis), University of York.
- Yang, H., Thompson, C., Hamm, R. M., Bland, M., & Foster, A. (2013). The effect of improving task representativeness on capturing nurses' risk assessment judgements: A comparison of written case simulations and physical simulations. *BMC Medical Informatics and Decision Making*, 13, 62. doi:10.1186/1472-6947-13-62

**Modeling and Aiding Intuition:
In the Twilight between the Conscious and the Unconscious**

Ulrich Hoffrage & Julian N. Marewski
University of Lausanne, Switzerland

Contact: ulrich.hoffrage@unil.ch, julian.marewski@unil.ch

Recently, the *Journal of Applied Research in Memory and Cognition* published a special issue entitled “Modeling and Aiding Intuition in Organizational Decision Making” (edited by Marewski & Hoffrage, 2015). Many of the 17 articles assembled there were closely related to various issues Egon Brunswik covered in his writings. The article that is probably closest to the work of Brunswik is the one of Ken Hammond (2015), on “Causality vs generality: Judgment and decision making struggles to become a scientific discipline”—presumably his last one. In this article, Ken focusses on representative design. He takes the reader back to the 1940ies and reports how Brunswik struggled to promote it as a methodological imperative, and what the reception of representative design was in the subsequent decades, until today (see also Dhimi, Hertwig, & Hoffrage, 2004). To reiterate, we “feel extremely glad and honored that we can include Ken’s article in the present special issue, and we believe he would be delighted as well if he could see the context (i.e., the other papers) in which his essay appears” (Hoffrage & Marewski, 2015, p. 155).

In our introduction to that special issue, we point out that much, if not most, of our judgment and decision making can be located between two poles: analysis and intuition (Hoffrage & Marewski, 2015). Obviously, this view is not new. It is the essence of Brunswik’s (1952) notion of *quasirationality*, which has been further developed in Hammond’s *Cognitive Continuum Theory* (Hammond, 1996, 2010; see also Dhimi & Thomson, 2012). Simon (1987) expressed the idea as follows: “intuition is not a process that operates independently of analysis; rather, the two processes are essential complementary components of effective decision-making systems” (p. 61). In our introduction, in which we referred to intuition as the Lady in Black, we adopted not only a broad perspective but also used an unusual writing style with many metaphors and dichotomies, such as day and night, light and dark, conscious and unconscious. Consistent with Brunswik’s notion of quasirationality, we concluded our historical and conceptual part as follows: “We ... met our dear Lady in Black in various forms. We met her as inspirations, possibly from sources that some might locate outside ourselves. We met her as animalistic instincts that come in form of intuitive statistics. We met her as statistical inferences that can, for instance, be described as stemming from fast-and-frugal heuristics. And finally we met her as insights that we may have when we engage in problem solving or restructuring. The common denominator is this: In each of these forms, something emerges from the dark side within us. What comes out of our darkness interacts with conscious awareness, it becomes an object that can be scrutinized and analyzed, and we may wonder “Where did this idea, this hunch, this gut feeling come from?” The inner light – thinking, reasoning, and rationality – that the eminent figures of the Enlightenment lifted on the throne which was, before, occupied by religious content, shines into the darkness and attempts to understand. We cannot

tell exactly where this inner light comes from and why it disappears while sleeping. ... But we feel that both sides, the inner day and the inner night, belong to us. Enlightenment and Romanticism have been eras in history, but they are more: What was driving their representatives is constituting each of us. Subconscious inspirations, instincts, inferences and insights are not intuitions, but once they leave the dark side within us and we become aware of them, they turn into intuitions. Adopting Goethe's theory of colors, intuitions are like colors that emerge in turbid media, in the sphere of twilight, where light and darkness meet and interact with each other. Intuitions are messengers that enter our all-joyous days, and they remind us of our own origin: "The mysterious night side" (Hoffrage & Marewski, 2015, p. 154).

Even more recently, the Journal of Applied Research in Memory and Cognition published a set of seven commentaries (edited by Hoffrage, Marewski, & Fisher, 2016) on various aspects and articles of the special issue. In the remainder, we provide a very brief overview of these commentaries. We grouped them into three categories, starting with those that questioned the nature of intuition and its relationship to analysis.

1. *Intuition seen in different dimensions and domains.* While some of the authors of the special issue focus on single (and different) aspects of intuition, such as unconscious decision-making, holistic decision-making, or effortless decision-making, others see multiple dimensions of intuition. Amit, Rusou, and Arieli (2016) contrast such uni-dimensional and multi-dimensional conceptualizations of intuition and review research that tested whether there is an empirical basis to separate various dimensions of intuition. The other two articles in this category address the question whether and when people favor intuition over analysis. Olds and Link (2016) build on Pachur and Spaar's (2015) findings, namely that people's preferences for one or the other style are domain-specific, and discuss two possible sets of explanations for such domain-specific preferences: objective characteristics of the domain itself (e.g., predictability of events), and individual differences (e.g., with respect to expertise). Szaszi (2016) raises a conceptual issue: Is people's tendency to adopt an intuitive decision style the same as their preference for that style? He issues a warning that these two concepts should not be used interchangeably.

2. *Intuition reflected in clashes and conflicts.* Kurt Lewin once remarked that there is nothing more practical than a good theory. Yet theory and practice do not always find each other, and sometimes, the two even conflict. Gore and Conway (2016) discuss the contrast of intuition and analysis in the fuzzy and uncertain world of organizations, that is, in practice. They call for a hybrid approach that combines intuition and analysis, and discuss (a) ways to build bridges between academia and practice and (b) how research in general and intuition research in particular can have greater impact on policy and practitioners. Frey, Neys, and Bago (2016) also focus on conflict, albeit not on the conflict between theory and practice, but on conflicts between intuitive and analytic forms of thinking: What happens when individuals detect that intuitive processes produce outcomes that violate normative standards?

3. *Intuition cast into the language of mathematics and computer code.* In the cognitive and decision sciences, various tools have been developed to model how people arrive at their judgments. The primary tool in the Neo-Brunswikian tradition is judgment analysis, but there are many others, including cognitive architectures such as ACT-R. Larue and Juvina (2016) discuss how ACT-R could be used to implement

the tri-partitive framework of Stanovich (2009), according to which the human mind can be decomposed functionally into an autonomous mind, an algorithmic mind, and a reflective mind. Larue and Juvina propose that the algorithmic mind, for instance, might be modeled in terms of ACT-R's memory mechanisms with its underlying mathematical equations. Finally, Bear and Rand (2016) demonstrate how computer simulations can be used as means towards building theories on the likely origins of intuition. They demonstrate that formal models inspired by evolutionary game theory can help explain why intuitions might have evolved even though agents who process information more thoroughly outperform "intuitive" agents.

Conclusion

"Intuition, so we realized in the course of almost four years of editorial work, continues to be a beautiful, elusive subject of scientific inquiry that seems to shy away from the light of analysis and that behaves like an inkblot test or a chameleon – different people construe intuition in different ways (...). Much can be learned about intuition through scientific definitions, equations, computer simulations, or experiments. At the same time, much would also be missed if one relied exclusively on such approaches (...). Intuition comes as an inner feeling, as a messenger from our unconscious, as something undefinable, but yet illuminating and powerful. Gestalt-inducing media, such as pictures, poetry, or other artworks, can transport that insight and get across some flavour of intuition's mystery (...): A breeze of romantic longing for something that is hidden and occult, yet, at the same time intimately related to us and our own origin" (Marewski, Hoffrage, & Fisher, 2016, p. 321).

References:

- Amit, A., Rusou, Z., & Arieli, S. (2016). Uni-dimensional or multi-dimensional? Empirical evidence for distinguishing between different aspects of intuition. *Journal of Applied Research in Memory and Cognition*, 5, 322–324. <http://dx.doi.org/10.1016/j.jarmac.2016.06.002>
- Bear, A., & Rand, D. G. (2016). Modeling intuition's origins. *Journal of Applied Research in Memory and Cognition*, 5, 341–344. <http://dx.doi.org/10.1016/j.jarmac.2016.06.003>
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago, IL: University of Chicago Press.
- Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, 130(6), 959–988. <http://dx.doi.org/10.1037/0033-2909.130.6.959>
- Dhami, M. K., & Thomson, M. E. (2012). On the relevance of Cognitive Continuum Theory and quasirationality for understanding management judgment and decision making. *European Management Journal*, 30(4), 316–326. [doi.org/10.1016/j.emj.2012.02.002](http://dx.doi.org/10.1016/j.emj.2012.02.002)
- Frey, D., de Neys, W., & Bago, B. (2016). The jury of intuition: Conflict detection and intuitive processing. *Journal of Applied Research in Memory and Cognition*, 5, 335–337. <http://dx.doi.org/10.1016/j.jarmac.2016.06.004>
- Gore, J., & Conway, G. E. (2016). Modeling and aiding intuition in organizational decision making: A call for bridging academia and practice. *Journal of Applied Research in Memory and Cognition*, 5, 331–334. <http://dx.doi.org/10.1016/j.jarmac.2016.08.002>
- Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford, England: Oxford University Press.
- Hammond, K. R. (2010). Intuition, No!... Quasirationality, Yes!. *Psychological Inquiry*, 21(4), 327–337.
- Hammond, K. R. (2015). Causality vs generality: Judgment and decision making struggles to become a scientific discipline. *Journal of Applied Research in Memory and Cognition*, 4(3), 176–179. <http://dx.doi.org/10.1016/j.jarmac.2015.07.002>
- Hoffrage, U., & Marewski, J. N. (2015). Unveiling the Lady in Black: Modeling and aiding intuition. *Journal of Applied Research in Memory and Cognition*, 4(3), 145–163. [doi: 10.1016/j.jarmac.2015.08.001](http://dx.doi.org/10.1016/j.jarmac.2015.08.001)

- Hoffrage, U., Marewski, J. N., & Fisher, R. P. (2016). Modeling and aiding intuition in organizational decision making: Commentary section. *Journal of Applied Research in Memory and Cognition*, 5, 318–344.
- Larue, O., & Juvina, I. (2016). A call for unification of dual- and single-process accounts in cognitive models of intuition. *Journal of Applied Research in Memory and Cognition*, 5, 338–340. <http://dx.doi.org/10.1016/j.jarmac.2016.06.007>
- Marewski, J. N., & Hoffrage, U. (2015). Modeling and aiding intuition in organizational decision making. *Special Issue in Journal of Applied Research in Memory and Cognition*, 4, 145-311.
- Marewski, J. N., Hoffrage, U., & Fisher, R. P. (2016). Modeling and aiding intuition: Introduction to the commentary section. *Journal of Applied Research in Memory and Cognition*, 5, 318–321. doi:10.1016/j.jarmac.2016.08.001
- Olds, J. M., & Link, D. (2016). Unpacking decision domains – Commentary on "Domain-specific preferences for intuition and deliberation in decision making". *Journal of Applied Research in Memory and Cognition*, 5, 325–328. <http://dx.doi.org/10.1016/j.jarmac.2016.06.015>
- Pachur, T., & Spaar, M. (2015). Domain-specific preferences for intuition and deliberation in decision making. *Journal of Applied Research in Memory and Cognition*, 4(3), 303–311. <http://dx.doi.org/10.1016/j.jarmac.2015.07.006>
- Simon, H. A. (1987). Making management decisions: The role of intuition and emotion. *The Academy of Management Executive*, 1(1), 7–64.
- Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford: Oxford University Press.
- Szaszi, B. (2016). The role of expertise and preference behind individuals' tendency to use intuitive decision style. *Journal of Applied Research in Memory and Cognition*, 5, 329–330. <http://dx.doi.org/10.1016/j.jarmac.2016.06.006>

**Learning to Make Selection and Detection Decisions:
The Role of Task Uncertainty**

Jim R. Holzworth

University of Connecticut, US

Tom R. Stewart

Center for Policy Research, University at Albany, State University of New York, US

Jeryl L. Mumpower

Texas A&M University, US

Contact: jimholzworth@gmail.com

Effects of task uncertainty on threshold learning in a multiple cue decision task were examined under two types of feedback and three base rate conditions. In most such decision experiments, participants receive feedback after every trial (full feedback) with a single (usually .5) base rate. Our experiment explored conditional (decision-contingent) feedback, in a task representing a detection problem (passenger screening) in which the decision maker receives no feedback unless the decision is positive (e.g., search the passenger). We manipulated three base rates (.1, .5 and .8) and three levels of task uncertainty (low, moderate, and high, defined by criterion variance accounted for: .9, .7, and .5). Increased uncertainty made all dependent measures worse. Task uncertainty had detrimental effects on both judgment and decision making, and interacted with effects of feedback and base rate. Performance was best with full feedback, but after 300 learning trials the difference with conditional feedback was small. There may be no single unifying explanation for results of our base rate manipulation. Conditional feedback generally resulted in fewer positive decisions than full feedback, but not in the low (.1) base rate condition. Results provide partial support for constructivist encoding and for accuracy maximization with moderate and high base rates, but not with a low base rate. Our results reflect overconfidence when conditional feedback was given in moderate and high base rate conditions, and an exploratory strategy when base rate was low.

**A Hammond-Brunswik Orientated Evaluation:
The Success of Linear Bootstrapping Models as an Example**

Esther Kaufmann

University of Zurich, Switzerland

Werner W. Wittmann

University of Mannheim, Germany

Contact: esther.kaufmann@gmx.ch

In the following, we describe our recent meta-analysis of the success of bootstrapping models within the lens model approach (Kaufmann & Wittmann, 2016). In this paper, we evaluated the success of replacing or ‘bootstrapping’ human judges with decision-making models (e.g., equations). Both the topic of our work— lens model studies—as well as the methodology of our meta-analysis are strongly related to the Hammond-Brunswik tradition. With regards to methodology, we note that Hammond’s recommendation to focus on an idiographic approach also applies to meta-analytic research. Traditional meta-analysis is based on the aggregated results of multiple studies (e.g., the average success of bootstrapping in each study, aggregated across a sample of judges and a number of judgment tasks). An alternative to traditional meta-analysis is individual participant data (IPD) meta-analysis. In IPD meta-analysis, individual-level data from multiple studies is pooled together and analyzed directly (for additional information, we refer to Stewart et al., 2015, p. 1657). Only IPD meta-analysis avoids the potential for several aggregation biases (e.g., ecological fallacy, Robinson, 1950; Simpson-Paradox, Simpson, 1951, see Kaufmann, Reips, & Maag Merki, 2016), which may skew the results of traditional meta-analyses. Hence, the goal of our paper was to use an IPD meta-analytic approach to evaluate the success of bootstrapping models. As another unique feature of our meta-analysis, we also considered whether the success of bootstrapping might depend on a) the decision domain (e.g., education or medicine), b) judges’ expertise level (novice vs. expert) within domains, and c) the evaluation criteria (subjective, test, objective).

We conducted a meta-analysis of 35 studies with 1,110 bootstrapping models, 532 experts / 578 novices judging 52 tasks across five decision domains. In line with Brunswik’s and also Hammond’s recommendation (see Dhami, 2015), we started our analysis by following an idiographic approach. Specifically, we plotted the judgment accuracy of single judges. However, our database revealed that individual-level data was available for only about one third of the total sample of bootstrapping models (365 out of 1,110) and about half of the sample of judgment tasks (28 out of 52). After our analysis of the individual-level data, we then conducted a meta-analysis based on the success of bootstrapping aggregated at the task-level, and finally a meta-analysis based on the success of bootstrapping aggregated at the study-level.

Bootstrapping models are thought to be especially useful when making important—and often ambiguous—decisions, such as reaching a medical diagnosis or choosing a candidate for a particular job (Swets, Dawes, & Monahan, 2000). Indeed, the results of our stepwise analysis indicated that bootstrapping is associated with

slightly more accurate judgment relative to human judgment across all decision domains. Moreover, our results also indicated that both novices and experts would profit from using formal decision-making models, especially when there is an objective evaluation criterion. The higher success of bootstrapping is meaningful particularly in high-risk decision-making domains like medical science, in which even small increases in decision accuracy could lead to many saved lives. In sum, our results support the conclusion that formal models to guide and support decisions should be developed, especially when the cost of inaccurate decisions is high. We refer interested readers to Kaufmann and Wittmann (2016) for more details about the study.

This paper extends our previous meta-analysis on lens model components (see Kaufmann, Reips, & Wittmann, 2013), focusing on linear bootstrapping models. In addition to revealing the greater success of bootstrapping models, our analysis also revealed that authors of bootstrapping studies seldom reported individual-level data. Individual-level data are seldom reported in other research fields, too, which may be one reason why IPD meta-analyses are so rarely conducted. An idiographic meta-analytic approach avoids potential aggregation biases and could also reveal if the results of previous traditional meta-analyses were erroneously based on ecological fallacies or Simpson's paradox. We therefore echo Hammond's and Brunswik's plea for researchers to report individual-level data.

References:

- Dhami, M. K. (2015). Ken and I. *Brunswik Society Newsletter*, 30, 10–11.
- Kaufmann, E., Reips, U.-D., & Maag Merki, K. (2016). Avoiding methodological biases in meta-analysis: Use of online versus offline Individual Participant Data (IPD) in educational psychology. *Special Issue. Hotspots in Psychology 2016. Zeitschrift für Psychologie*, 224, 157–167.
- Kaufmann, E., Reips, U.-D., & Wittmann, W. W. (2013). A critical meta-analysis of lens model studies in human judgment and decision-making. *PLoS ONE* 8(12): e83528. doi:10.1371/journal.pone.0083528
- Kaufmann, E., & Wittmann W. W. (2016). The success of linear bootstrapping models: Decision domain-, expertise-, and criterion-specific meta-analysis. *PLoS ONE* 11(6): e0157914. doi:10.1371/journal.pone.0157914
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(2), 351–357.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*, 13, 238–241.
- Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G., & Tierney, J. F. (2015). Preferred reporting items for a systematic review and meta-analyses of individual participant data: The PRISMA-IPD statement. *Journal of the American Medical Association (JAMA)*, 313(16), 1657–1665. doi:10.1001/jama.2015.3656
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1–26. doi:10.1111/1529-1006.001

**Using Brunswik's Lens Model to Nail
the (Unstructured) Interview**

Edgar E. Kausel

*Pontificia Universidad Católica de Chile,
School of Management, Chile*

Contact: ekausel@uc.cl

In a paper that we recently published in *Organizational Behavior and Human Decision Processes*, we used the Lens Model to study how the way hiring managers weight different cues (in particular, information from unstructured interviews) can decrease their accuracy levels while increase their confidence (boosting their overconfidence).

Previous research in personnel selection had shown three things. First, General Mental Ability (GMA) tests—whether we like it or not—have decent environmental validity in predicting employee job performance (though far from perfect, of course). Second, personality tests (in particular, conscientiousness tests), have some incremental validity over GMA tests. Third, unstructured interviews have low validity in predicting performance; and when used in combination with standardized tests (GMA and conscientiousness tests), they add zero in predicting employee performance. This, in Brunswikian terms, pertains to the left side of the lens model applied to personnel selection. In addition, regarding the right side of the lens model, there is some evidence suggesting that managers weight highly the unstructured interview, and much less GMA tests.

Using these previous findings and based on the lens model, what we hypothesized and found is that actual managers who are presented with information only from standardized tests did better in predicting employee performance than those who were presented with this same information but also with information from unstructured interviews.

In other words, while previous research had shown that unstructured interviews have low (environmental) validity, our study suggests that in some circumstances (that is, in the presence of environmentally valid cues) adding unstructured interview information can hurt hiring decisions.

I should also report that Brunswik's lens model is particularly useful in teaching personnel selection in human resource management courses, because it can highlight how cues that people utilize may not be particularly valid in the environment. It's also useful when teaching about discrimination in this context—for example, managers may weight race as a cue while it may have low or zero environmental validity.

Finally, I should note that the Brunswik model is one of the building blocks of a course I teach on Rationality and Decision Making. The “forecasting and predictions” section is based on Brunswik's, Hammond's, and Hogarth and Karelaia's conceptual papers and meta-analyses. It's challenging to teach but very rewarding in the end.

(I'll always be grateful to my dissertation advisor and mentor, Terry Connolly, for using the lens model to understand JDM in his course back in my doctoral years at the University of Arizona.)

References:

- Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, 137, 27–44. doi:10.1016/j.obhdp.2016.07.005
- Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, 134(3), 404–426. doi:10.1037/0033-2909.134.3.404

**Brunswik's Lens Model Approach
in Agent-based Simulations on Crowd Wisdom**

Marc Keuschnigg
*Institute for Analytical Sociology,
Linköping University, Sweden*

Christian Ganser
*Department of Sociology,
LMU Munich, Germany*

Contact: marc.keuschnigg@liu.se / christian.ganser@lmu.de

Firms and forecasters frequently combine multiple judgments (Armstrong, 2001; Bonabeau, 2009) to benefit from the “wisdom-of-crowd effect” (Surowiecki, 2004): When predicting an unknown outcome, the average over a collection of independent judgments represents the truth more closely than the typical individual guess. With increasing group size and variation, individual judgments are likely to fall on both sides of the truth, permitting aggregation to cancel out contradictory biases. Consequently, heterogeneous groups almost always outperform samples of homogeneous experts—even when the crowd consists of error-prone individuals. Hence, prior research claimed that in collective decision-making “diversity trumps ability” (Hong & Page, 2004; Page, 2007). Our research, recently published in *Management Science* (Keuschnigg & Ganser, 2016), demonstrates that this proposition is not universal, but relies on the averaging principle to generate group decisions. This is particularly irritating, because in practice averaging is rarely used as an aggregation rule. Instead, most group decisions in social life rest on voting.

We use agent-based simulations to quantify the benefits from heterogeneity and expertise across aggregation rules. We rely on Brunswik's (1955) lens model approach as a plausible representation of agents' behavior within a variable task environment: The environment provides a criterion (e.g., the future price of a stock) and a set of probabilistically related cues (e.g., a firm's profits, innovativeness, and market position). The task entails individual selection of the superior alternative. Agents receive multiple cues which they combine to provide an individual estimate of the criterion. Agents differ in how precisely they perceive cues and how apt they are in combining them (our manipulation of ability). We randomly pick agents to create nominal groups of different size. Group members feature different judgment errors (our manipulation of diversity). We then apply averaging and plurality vote to aggregate members' judgments into a single group solution.

Although the lens model is unlikely to provide full understanding of human judgment, it is highly valuable in directly comparing aggregation rules (see also Hastie & Kameda, 2005) and evaluating their dependency on specific group characteristics. Brunswik's framework allows the taking into account of a wide range of parameter values including the overall task environment, individual characteristics, and group features. We set parameter ranges at plausible real-world intervals (Karelaia &

Hogarth, 2008 provide reference values). With this calibration of our simulation's behavioral core, we closely reproduce findings on real human judgment.

Our results replicate the “diversity trumps ability” proposition for large groups. Samples of heterogeneous agents outperform same-sized homogeneous teams of high ability. This is true for both averaging and voting. Diversity, however, is not universally important. In small groups, its influence on collective accuracy strongly depends on the aggregation rule employed: Whereas diversity remains crucial under averaging, in the case of voting agents' individual abilities determine collective accuracy. Depending on the simulation's calibration, the threshold up to which ability dominates collective accuracy lies at around 14–20 members. Hence, voting in small groups—which is ubiquitous in both public administration and private firms—benefits more from individual expertise than from diversity. Of course, our generalizations are limited to groups free from social interaction and are thus only a necessary benchmark in evaluating decision making in real human groups. Still, our results highlight that one should not readily apply strategies which solely rest on maximizing diversity to improve electorates' accuracy.

References:

- Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Boston, MA: Kluwer.
- Bonabeau, E. (2009). Decisions 2.0: The power of collective intelligence. *Sloan Management Review*, 50(2), 45–52.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62(3), 193–217.
- Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, 112(2), 494–508.
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the Academy of Science of the United States of America*, 101(46), 16385–16389.
- Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, 134(3), 404–426.
- Keuschnigg, M., & Ganser, C. (2016). Crowd wisdom relies on agents' ability in small groups with a voting aggregation rule. *Management Science*, Article in Advance. doi:10.1287/mnsc.2015.2364.
- Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, NY: Doubleday Books.

Towards an Ecologically Valid Model of Spontaneous Stereotypes about Groups

Alex Koch

University of Cologne, Germany

Roland Imhoff

Johannes Gutenberg University, Mainz, Germany

Contact: alex.koch@uni-koeln.de

Stereotyping means to perceive groups as scoring low to high on certain dimensions. Perceivers use groups' variance on these proximal, tangible dimensions (i.e., stereotypes) to predict group members' variance on distal, intangible dimensions that are relevant to the self. For example, based on perceiving librarians and comedians as low and high on outgoingness, respectively, party animal John predicts that comedian Paul goes partying more often than librarians George and Ringo. In terms of Brunswik's lens model (1952), partying predictions may depend more on outgoingness compared to funniness stereotypes, and people may to some extent use other stereotypes, too, such as groups' perceived attractiveness. Regardless of the extent to which people use stereotypes, their utility may be low to high. For example, outgoingness may, and funniness and attractiveness stereotypes may not, predict actual partying.

Brunswik (1952) argued that stereotype use is a function of stereotype utility. So, having learned that outgoingness stereotypes best predict actual partying, people's partying predictions will depend more on their outgoingness compared to other stereotypes. However, if the groups in a social context do not vary in perceived outgoingness, outgoingness stereotypes do not predict actual partying in that social context, and thus people will revert to other stereotypes. People will also revert to other stereotypes if they temporarily cannot access their outgoingness stereotypes. Generally speaking, as accessible stereotype variance is a necessary (but not sufficient) condition for stereotype utility, from the lens model it follows that stereotype use is a function of accessible stereotype variance. Thus, to draw ecologically valid inferences about people's stereotype use, lab research on stereotype use must preserve the accessible stereotype variance in people's social environment. In other words, research on stereotype use must comply with Brunswik's (1955, 1956) call for representative design.

Unfortunately, previous research on stereotype use did not make sure to preserve the accessible stereotype variance in people's social environment. In more or less all studies that support the model that people mainly use warmth and competence stereotypes (e.g., Cuddy, Fiske, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002), participants rated groups on only just stereotypic warmth and competence. As this design unnaturally denies participants access to group variance on stereotypes other than warmth and competence, studies that used this design do not provide insight into the stereotypes that people use in their social environment. What is needed is a design that allows participants to use any desired stereotype(s).

A design in which participants rate the similarity of groups does the trick. Similarity needs to be construed with respect to one or another stereotype before it can be rated. For example, the similarity of politicians and doctors can be construed and rated with respect to their warmth, competence, or any other desired stereotype. The stereotypes that participants spontaneously use to rate the similarity of groups can subsequently be figured out based on ratings of the groups on candidate stereotypes. Importantly, these spontaneously used stereotypes only reflect the stereotypes that people use in their social environment if participants rate the similarity of groups that are representative of their social environment.

To figure out the stereotypes that people use in their social environment, we thus started with putting together a representative sample of groups. We instructed participants to name the groups that they think society is divided into. Every group named by at least a tenth of participants we included into our arguably representative sample. Next, we instructed other participants to drag and drop the groups back and forth on the screen, placing more similar groups closer together. This spatial arrangement method (SpAM; Koch, Alves, Krüger, & Unkelbach, 2016) is highly efficient (Hout, Goldinger, & Ferguson, 2013), as rearranging an item simultaneously readjusts the similarity proximities between that item and all other items. Next, we scaled the mean similarity proximities in a statistically well-fitting 2D space. In this space the groups were modeled by points, and the similarity of the groups was modeled by their proximity. To reveal the stereotypes that participants had spontaneously used to rate the similarity of the groups, we predicted the groups' mean ratings on candidate stereotypes based on the groups' coordinates in the 2D similarity space.

Results showed that participants had rated the groups' similarity based on their stereotypic agency / socioeconomic success (A; ~hierarchy; not the same as competence!) and conservative-progressive beliefs (B; ~ideology). Surprisingly, we found no evidence for the use of warmth / communion (C; ~likeability) stereotypes. However, groups' stereotypic C could be modeled as averageness on A and B, resulting in an ABC model of spontaneous stereotypes about groups. So, consistent with Aristotle's notion of a virtuous mean flanked by the vices of insufficiency and excess, groups seen as more average on A and B were seen as higher on C (Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016). The ABC model held true across 20+ countries (Imhoff & Koch, 2016) and both groups and states as targets of stereotypes (Koch, Kervyn, Kervyn, & Imhoff, 2016).

The utility of the ABC model increases with evidence for effects of spontaneous A and B stereotypes on intergroup emotions and behavior (intentions). In new experiments people's spontaneous beliefs stereotypes determined their intentions to cooperate in economic games. Participants were more willing to team up with members of stereotypically progressive groups (e.g., musicians and vegans) in game conditions that required choosing risky options or alternative options. In contrast, participants preferred members of conservative groups (e.g., religious people and senior citizens) in game conditions that required choosing safe options or the status quo option. Further, participants preferred members of stereotypically agentic compared to unagentic groups regardless of game condition.

In sum, complying with Brunswik's (1952; 1955; 1956) call for representative design, we developed an ecologically valid ABC model (agency / socioeconomic

success, conservative-progressive beliefs, and communion) of spontaneous stereotypes about groups.

References:

- Brunswick, E. (1952). *The conceptual framework of psychology*. *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.
- Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62(3), 193–217.
- Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92(4), 631–648.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902.
- Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, 142, 256–281.
- Imhoff, R., & Koch, A. (in press). How orthogonal are the Big Two of social perception? On the curvilinear relationship between agency and communion. *Perspectives on Psychological Science*.
- Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(8), 1171–1192.
- Koch, A., & Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency / socio-economic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology*, 110, 675–709.
- Koch, A., Kervyn, N., Kervyn, A., & Imhoff, R. (under review). *Lessons from U.S. citizens' map of the U.S. states: Interstate similarity in stereotypic prosperity and ideology predicts interstate liking*.

**Modeling Managerial Decision Making:
A Brunswikian Perspective**

Martin H. Kunc

Warwick Business School,
University of Warwick, UK

Contact: Martin.Kunc@wbs.ac.uk

Studies on managerial decision making taking a Brunswikian research approach are very infrequent in mainstream management literature. This essay discusses some of my personal experience and reflections on using a Brunswikian approach to understand and model managerial decision making as discussed in a recent book chapter (Kunc, 2016).

Basic process of judgmental accuracy. Consider for a moment the task of a manager who is making decisions related to a business. The manager's knowledge about the business is updated as he receives information over time from different sources, e.g., financial reports, meetings with other managers, evaluation of products, etc., which is processed using different weights (e.g., a financially minded manager may put more emphasis on financial reports than on oral reports from other managers). One important aspect that is missing in studies on managerial decision making is the process taken by the manager to adapt to the task environment, e.g., the development of judgmental accuracy in terms of selection of cues and change of weights. An important corollary of this situation is the importance of a processual approach to managerial decision making where it is necessary to evaluate how close manager's perception of a business is to the business's real components and how the accuracy changes over time as it changes cues and weights. In one of my studies on the impact of performance measurement systems, I found that a widely known performance measurement system only captured on average 50% of the variables employed to understand a business and most of the time the weights were incorrect (Kunc, 2008).

Information selection and its influence on decision making. The subjective interpretation of the business information is subject to an internal feedback process based on the dissatisfaction between the level of judgment and the performance obtained. People as adaptive systems are dominated by goal-seeking feedback process. Thus, the determination of the adjustment in each cue is determined by the level of dissatisfaction. One important consequence is the physical impossibility to update judgment before evidence is presented, e.g., the success of a strategy, which implies changes in a business, can only be measured after its implementation. Thus, managerial decision making accuracy is improved over time once the subject is able to interpret the evidence presented. However, interpretation can be heterogeneous due to structural differences across managers that affect cues and weights. In a simulation game played with more than 200 students, I found there were four different interpretations of the game (actions and results) which led to different decision making processes. The pattern of decisions only changed when there was enough evidence about poor performance, usually at the end of the game (Kunc & Morecroft, 2010).

Final considerations. Functionalism has illuminated some areas of research related to modeling managerial decision making. For example, it demonstrated that learning from outcome feedback is slow and limited (because it provides only few cues) leading to the development of cognitive feedback (a much richer set of cues) (Todd & Hammond, 1965). Cognitive feedback is at the core of behavioral experimentation related to misperceptions of feedback processes in managerial decision making (Kunc, 2012).

However, there is additional need for the Brunswik model to illuminate research in managerial decision making. For example, experimentation with managerial decision making should follow representative designs rather than systematic design (Goldstein, 2004). Representative design implies the design of experiments should reflect the natural environment (stimuli and conditions) of the managers to reveal issues in judgement accuracy. Therefore, mainstream models of managerial decision making need to replicate when managers are accurate, ecological validity, without portraying managers as perfectly rational decision makers.

References:

- Goldstein, W. M. (2004). Social judgment theory: Applying and extending Brunswik's probabilistic functionalism. In D. J. Koehler & H. Harvey (Eds.), *Blackwell handbook of judgment and decision-making* (pp. 37–61). Malden, MA: Blackwell Publishing.
- Kunc, M. (2008). Using systems thinking to enhance strategy maps. *Management Decision*, 46(5), 761–778.
- Kunc, M. (2012). Teaching strategic thinking using system dynamics: Lessons from a strategic development course. *System Dynamics Review* 28, 28–45.
- Kunc, M. (2016). Modeling behavioral decision making: Creation and representation of judgment. In M. Kunc, J. Malpass & L. Withe (Eds.). *Behavioral operational research. Theory, methodology and practice* (pp. 161–175). UK: Palgrave Macmillan UK.
- Kunc, M., & Morecroft, J. (2010). Managerial decision-making and firm performance under a resource-based paradigm. *Strategic Management Journal*, 31(11), 1164–1182.
- Todd, F. J., & Hammond, K. R. (1965). Differential feedback in two multiple-cue probability learning tasks. *Behavioral Science*, 10, 429–435.

**How the Risk-Reward Relationship Shapes
Decisions Under Risk and Uncertainty**

Christina Leuker, Timothy J. Pleskac, Thorsten Pachur, & Ralph Hertwig
Max Planck Institute for Human Development, Berlin, Germany

Contact: leuker@mpib-berlin.mpg.de

“Nothing ventured, nothing gained.” Financial advisors and lay people alike know that they sometimes need to take risks in order to obtain the big rewards they desire. In many real-world ecologies, risks and rewards, or payoffs and probabilities, are inversely related. Domains range from monetary gambles like the state lottery, roulette or the horse track to nonmonetary gambles such as the journal submission of scientific papers. Winning the jackpot in the lottery is not as likely as winning smaller amounts. Publishing in Nature is less likely than publishing in lower impact journals.

Different mechanisms give rise to this negative risk-reward relationship. For example, in man-made markets, consumer choices may remove dominated options: No (or maybe few) gambler(s) would bid on an option that promises both a very low payoff and a very low probability. No researcher would submit to a journal that has a very low acceptance rate and a low impact factor (for a detailed ecological analysis see Pleskac & Hertwig, 2014).

However, there seems to be a domain in which the negative risk-reward relationship is not present, and that is in many laboratory tasks investigating risky decision-making. Participants are often asked to choose between lottery-like gambles of the form “ p chance of winning x , otherwise nothing”. Here, risks and rewards are typically uncorrelated, or factorially combined (Pleskac & Hertwig, 2014). This orthogonalization makes sense because researchers want to extract the unique influence of payoffs and probabilities. But at the same time, this does not fit the ecological structure that people typically encounter outside the lab. According to an adaptive view of cognition, people select decision strategies that match the structure of the environment (Brunswik, 1943; Payne, Bettman, & Johnson, 1993; Simon, 1956). One way to exploit the risk-reward relationship, for example, is to use the reward magnitude as a signal for underlying probabilities, when probabilities are not explicitly stated.

Do participants’ choices differ systematically in more ecological, negative risk-reward environments versus uncorrelated, or even positive risk-reward environments? To test this, we exposed 152 participants to different risk-reward relationships in the lab, using monetary gambles of the form “ p chance of winning x , otherwise nothing”. Between-subjects, risks and rewards were negatively correlated, positively correlated or uncorrelated. Participants either chose between two gambles, or indicated their willingness to sell for single gambles.

We measured the influence of the different risk-reward ecologies (1) on a common set of gambles that appeared in each of the three conditions, in (2) an uncertainty task, in which participants were choosing between an uncertain option and

a sure thing, (3) an estimation task, in which participants made probability judgments based on reward magnitudes, and (4) a recognition task.

Participants were sensitive to the different risk-reward structures in both studies, notably without any external reinforcement in the exposure phase. When pricing single gambles, participants in the negative (but not positive or uncorrelated) condition performed much faster as trials progressed: This condition fits the priors people bring to the lab. In decisions under uncertainty, participants in the negative condition preferred the uncertain option when (both) payoffs were small – but not when payoffs were large. This is consistent with participants inferring probabilities from payoff magnitudes. We confirmed this in an explicit probability estimation task. Participants saw varying reward magnitudes and were asked to estimate how likely it was to win each of them, given their experiences in our experiment. Participants in the correlated condition were able to do this and estimated a positive/negative risk-reward association, given their condition.

Interestingly, some participants from the positive condition estimated a negative risk-reward relationship despite their experiences during the exposure phase. In the recognition task, participants rejected gambles that did not fit the structure they had experienced in the exposure phase (targets and lures). This suggests that participants were sensitive to the overall structure (but did not encode specific exemplars).

Pleskac and Hertwig (2014) showed that the ‘default’ risk-reward structure that people encounter in the world is negative. The current studies support this, given (for instance) faster pricing of gambles when this relationship is maintained. We also show that decision makers adapt to different risk-reward structures, and subsequently use them in decisions under risk and uncertainty. They infer (subjective) probabilities when they are not explicitly stated, and evaluate single options as drawn from a larger structure.

Key reference:

Leuker, C., Pleskac T. J, Pachur, T., & Hertwig, R. (in prep.). How risk-reward relationships shape decisions under risk and uncertainty.

References:

- Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, 50(3), 255–272.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.
- Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, 143(5), 2000–2019.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138.

Clinical Judgments in Acute Medical Care

Izhak Nadler

Contact: itsiknadler@gmail.com

Clinical judgments aimed to determine the state of a patient and the interventions that follow can be described with Brunswik's Lens Model. Presumably, clinical judgments in general and judgment made in intensive care units need to be accurate in order to provide the appropriate treatment. In practice, there is no measure to indicate the illness severity of patients who require acute medical care and consequently with no well-established reference measure, accuracy cannot be determined.

As an exception to that generalization, a measure was introduced to signal the medical distress of newborns that require some level of support to initiate independent respiratory immediately after birth. Dr. V. Apgar used a simple mathematical formulation based on five clinical signs to compute the clinical state of the newborns and accordingly to determine the required support level (see Apgar, 1953).

Thus, Dr. Apgar proposed a judgment policy during the mid-1950s, most likely without being aware of Brunswik's ideas that were published during the same era. Since its introduction there have been debates about the validity of the Apgar Score, but it is still being used all over the world.

In our first study, clinicians who provide neonatal resuscitation during their routine practice viewed a compilation of 51 video clips that presented a neonate mannequin in a variety of clinical states immediately after birth (Nadler, Liley, & Sanderson, 2010). Each clinician individually provided an Apgar Score representing her/his judgment about the state of the mannequin as shown at the end of the clip. The scores were used to compute the judgment policy of each clinician; clinicians' interpretations of the situations were found to be highly consistent, and describable by each individual's judgment policy. The participants' assessments were accurate when compared to the Apgar Scores that were computed out of the clinical signs that the mannequin presented at the end of each clip. Consequently, the agreement level between the clinicians was high and significant.

A study with a similar apparatus was used to capture clinicians' judgment policies with respect to babies who are hospitalized in the Neonatal Intensive Care Unit (NICU) (Nadler, Globus, Pessach-Gelblum, Strauss, & Ziv, 2014, 2016). Most of these babies are hospitalized to allow their organs to develop following a preterm birth. Currently, there is no reference measure to indicate the clinical state of such patients when their clinical state deteriorates and clinical intervention is required. Similarly to the former study, most clinicians demonstrated a consistent interpretation probably as a result of a pre-established policy. However, the judgment policies varied remarkably, resulting in a low agreement between the clinicians' judgments. Accuracy could not be tested due to lack of a reference score.

There are three main messages emerging from the findings of the two studies:

- Clinicians could use judgment policies that were developed during practical experience to assess and to quantitatively express their judgments about the illness severity of a simulated patient.
- Clinical judgments made in acute healthcare context are not different from any other judgments made by individuals under complex situations which require prompt response.
- The existence of a common reference increased the coherence of judgments made by a group of individuals.

The method developed and implemented in these studies demonstrates how simulation and video recordings can be used as means to capture judgment policies in highly dynamic environments. With these policies in hand, it is possible to measure various performance aspects of the judges while they take an active part as team members in a simulated scenario (Nadler & Sanderson, 2011; Nadler, Sanderson, & Liley, 2011).

With minimal changes the same process can be used to test and measure clinicians' performance in actual practice. The main implementation challenges refer to medico-legal aspects associated with patient rights and ethical conduct.

Judgment Analysis studies can promote healthcare by providing objective evidence for clinicians' performance that mainly relates to cognitive and social aspects of their practice. Such evidence is necessary to explore and reduce the number of cases in which highly skilled individuals fail to perform up to their capabilities.

In tribute to Kenneth R. Hammond who passed away in May 2015, it was Hammond's paper 'Ecological Validity: Then and Now' (1998) that inspired my PhD research at its early stages, exposing me to E. Brunswik and to Judgment Analysis. Not less importantly, Ken's paper is an inspiring call for integrity that should be an integral part at every aspect of academic conduct.

References:

- Apgar, V. A. (1953). A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anesthesia and Analgesia*, 32(4), 260–267.
- Hammond, K. R. (1998). *Ecological validity: Then and now*. The Brunswik Society. Web Essay. Retrieved November 1, 2016 from <http://www.brunswik.org/notes/essay2.html>
- Nadler, I., Liley, H. G., & Sanderson, P. M. (2010). Clinicians can accurately assign Apgar scores to video recordings of simulated neonatal resuscitations. *Simulation in Healthcare*, 5(4), 204–212.
- Nadler, I., & Sanderson, P. M. (2011). Using Brunswik's probabilistic functionalism to test how clinicians make judgments in simulated neonatal resuscitation scenarios. *Proceedings of the 55th Annual Meeting of the Human Factors and Ergonomics Society*, 743–747.
- Nadler, I., Sanderson, P. M., & Liley, H. G. (2011). The accuracy of clinical assessments as a measure for teamwork effectiveness. *Simulation in Healthcare*, 6(5), 260–268. doi: 10.1177/1071181311551154
- Nadler, I., Globus, O., Pessach-Gelblum L., Strauss, Z., & Ziv, A. (2014). *Using video recorded neonatal intensive care unit scenarios to test agreement between clinicians' assessments*. Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting. Chicago, IL. 27-31 October 2014.
- Nadler, I., Globus, O., Pessach-Gelblum, L., Strauss, Z., Sela, R., & Ziv, A. (2016). Applying judgment analysis theory and methods to obtain an insight into clinical judgments: Implementation and findings with a simulated neonatal intensive care unit setup. *Simulation in Healthcare*, 11(3), 200–208.

**A Neuroergonomics Approach to Cognition
in Judgment and Decision Making**

Joseph K. Nuamah & Younho Seong
North Carolina A&T State University, US

Contact: jknuamah@aggies.ncat.edu

The Cognitive Continuum Theory (CCT), proposed by Hammond is a coalescing theory for the field of human judgment and decision making (Cooksey, 1996). Hammond based the CCT on five premises. The fifth premise, Structural Counterparts Premise (Hammond, 1980), which attempted to relate the cognitive continuum to brain function, was dropped. According to Cooksey (1996, pp. 25), "...the dropping of the Structural Counterparts premise may have been premature and should probably be reinstated to give impetus to research on specific brain correlates of judgment." What we seek to do in our proposed study is to give impetus to research in this area by investigating the neural correlates of intuitive and analytical decision making during human automation interaction.

Our approach is based on neuroergonomics, the study of brain and behavior at work (Parasuraman, 2003). Neuroergonomics merges knowledge and methods from neuroscience and ergonomics, and explains the neural bases of both physical and cognitive performance that are left unanswered with traditional ergonomic assessments (Mehta & Parasuraman, 2013). Psychophysiology allows for the use of physiological measurements to understand a human operator's behavior by non-invasively recording peripheral and central physiological changes while the human operator behaves under controlled conditions. Psychophysiological signals are attractive because they are continuously available and their collection does not interfere with the operator's job performance. Thus it should be possible to observe changes in a human operator's brain signals in order to examine his or her cognitive processes in response to intuition-inducing and analysis-inducing tasks in human automation interaction. Electroencephalography (EEG) is a psychophysiological technique for studying brain activation. EEG signals represent summed postsynaptic potentials of neurons firing, sampling by millisecond. Graphically, an EEG is a graph of the time varying voltage difference between an active electrode attached to the scalp and a reference electrode (Gevins & Smith, 2006).

In our proposed study, a Hybrid Lens Model (Seong et al., 2006) depicted in Figure 1 is developed to represent outputs from an automated decision aid (ADA) and a human operator's judgments. The state of the environment is presented via cues. These cues have probabilistic relationships with the state of the environment. The ADA estimates the state of the environment based on available information, utilizing a fuzzy logic algorithm. The ADA may not be completely capable of estimating the true state of the environment, and this may have an effect on the human operator's judgment performance. The human operator assesses the state of the environment based on information provided as well as recommendations of state estimates from the ADA. The human operator is at liberty to rely on the recommendations from the ADA.

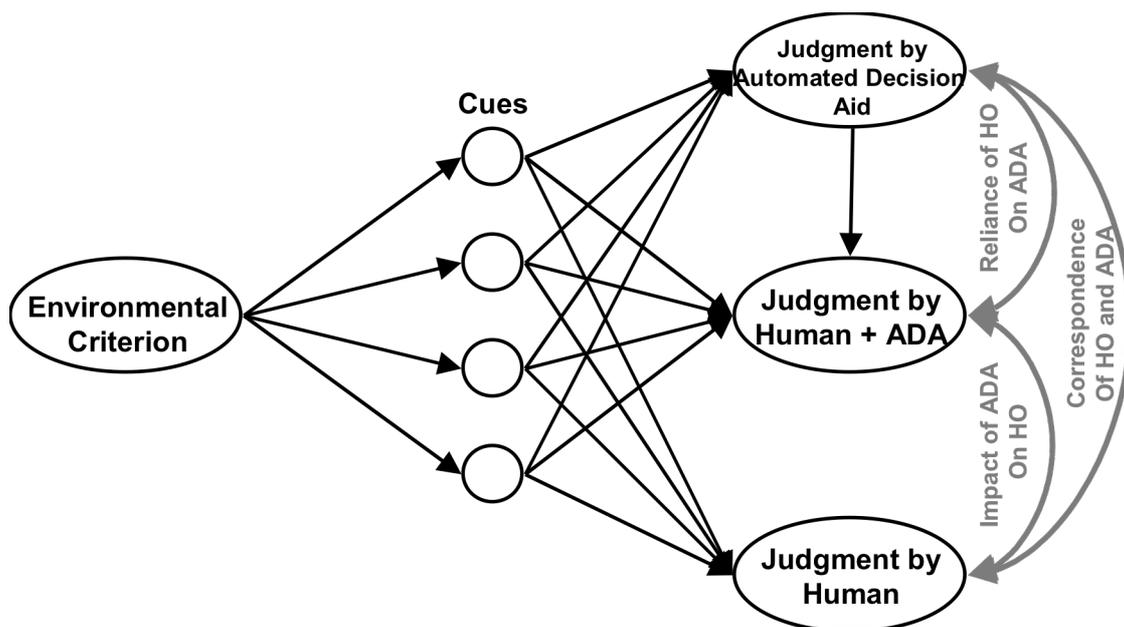


Figure 1. Hybrid n-system/hierarchical lens model. Three judgment agents (an automated decision aid (ADA), a human acting alone, and a human acting in concert with the automated aid) are shown. Note that the output from the ADA acts as an additional cue for the human + ADA system.

The ADA (Figure 2) in our study is designed to help human operators detect water quality anomalies in a drinking water distribution network. The system uses sensor sites installed through the distribution network to sample water at regular intervals. Each sensor site contains an array of four sensors. Appropriate sets of cues are used for constructing judgment profiles. A dichotomous judgment measure is employed for the human operator to judge whether or not a water profile is of good quality.

Three task conditions are presented to each participant - baseline (participant is instructed to rest, and not to do or focus on anything), intuition-inducing task (graphical representation of profile in an intuitive display), and analysis-inducing task (numerical representation of cues in a table format). For each task condition, EEG signals are recorded and analyzed offline.

For data analysis, logistic regression analysis will be performed on each side of the lens model to obtain predicted and residual value for each judgment record. For the two task conditions (intuition-inducing and analysis-inducing), recorded EEG signals will be analyzed for spectral and right-left hemisphere differences.

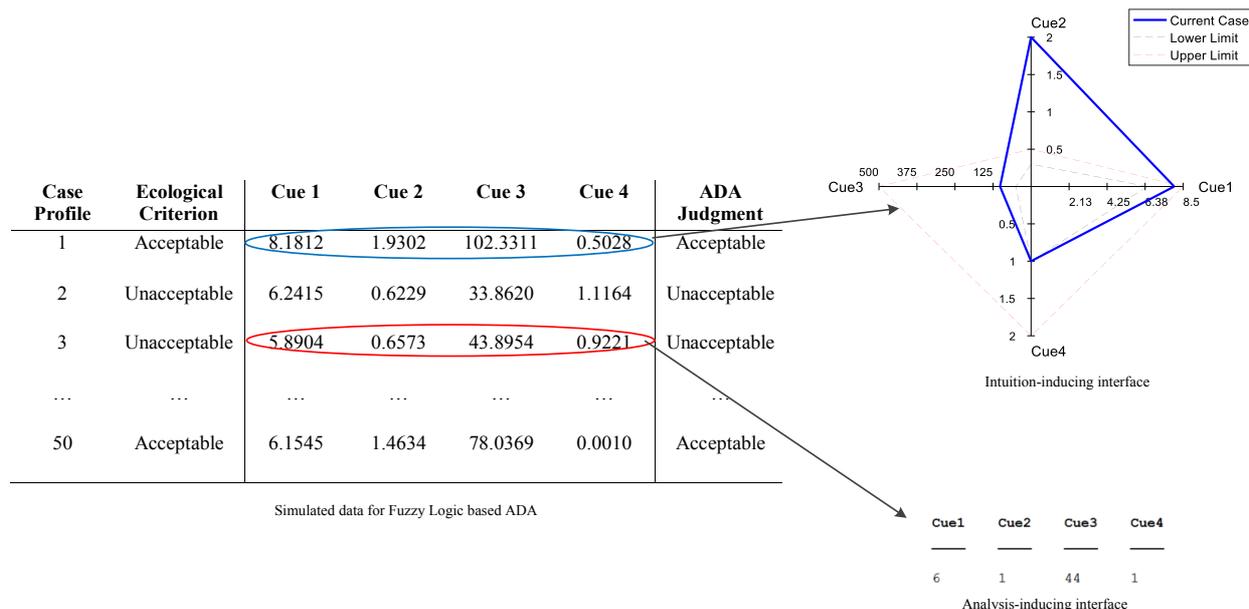


Figure 2. Fuzzy Logic Based ADA to induce Intuition and Analysis Cognition.

References:

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

Gevins, A., & Smith, M. E. (2006). Electroencephalography (EEG) in neuroergonomics. *Neuroergonomics: The brain at work*, 15–31.

Hammond, K. R. (1980). *The integration of research in judgment and decision theory*. Center for Research on Judgment and Policy. Boulder, CO: Institute of Behavioral Science, University of Colorado. Report No. 226.

Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: A review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*, 7, 889.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, 4(1-2), 5–20.

Seong, Y., Bisantz, A. M., & Gattie, G. J. (2006). Trust, automation, and feedback: An integrated approach. In A. Kirlik (Ed.), *Adaptive perspectives on human-technology interaction: Methods and models for cognitive engineering and human-computer interaction* (pp. 105–113). Oxford University Press.

**The Importance of Representative Design
in Human Factors Automation Research**

Richard Pak
Clemson University, US

Ericka Rovira
U.S. Military Academy Department of Behavioral Sciences & Leadership, US

Anne Collins McLaughlin
North Carolina State University, US

Contact: richpak@clemson.edu

Historically, technological automation was relegated to specific use cases with highly trained users (e.g., complex flight automation, process-control). However, users of all types are being exposed to more automation in many consumer-oriented contexts stressing the need to better understand the factors that influence their trust. The extent to which users trust automation is a key determinant of use and reliance (Lee & Moray, 1992, 1994; Muir & Moray, 1996).

It is assumed that the current body of trust in automation literature, carried out in one context with a specific set of users, will be generalizable to other contexts and users as long as the conceptual aspects of the systems being studied (e.g., reliability of automation) are identical. However, the majority of extant research on this topic, and generated theories and models, has typically focused on conducting the experiments under highly specialized domains such as aviation or industrial process control tasks and with highly trained populations such as military personnel or industrial workers (Hoff & Bashir, 2015).

As Brunswick was keenly aware, situations like this pose a serious threat to generalizability because while the experiments may have been precisely designed, the sampling of task situations, or domains of automation use, and user groups was limited (Brunswick, 1956). In an eloquent and clear essay, Hammond (1998) reiterated the importance of the concept of representative design for experimental psychologists but also clarified the common error of confusing it with the concept of “ecological validity.”

With representative design in mind, we (Pak, Rovira, McLaughlin, & Baldwin, 2016) recently examined how trust in automated technologies varied across different task situations or domains with a variety of population types. In contrast to research that examined trust in automation using the very narrow domains of industrial or military systems, we sampled domains that were likely to be encountered by the general population (e.g., consumer systems). We also deliberately sampled different user groups that varied in age but also came from different organizational cultures (civilian vs. military). We found that depending on the domain, age, and organizational-cultural background of the participant, important system-related variables such as reliability had different effects on trust. This finding was not originally anticipated from most models of trust as most existing research had been silent on these key variables.

The results not only highlighted the importance of the concept of representative design but also suggested caution in generalizing existing automation research to other domains or users.

References:

- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.
- Hammond, K. R. (1998). *Representative design*. The Brunswik Society. Web Essay. Retrieved November 1, 2016 from <http://brunswik.org/notes/essay3.html>
- Hoff, K. A., & Bashir, M. (2015). Trust in automation integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434. doi:10.1177/0018720814547570
- Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40(1), 153–184.
- Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270.
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429–460.
- Pak, R., Rovira, E., McLaughlin, A. C., & Baldwin, N. (2016). Does the domain of technology impact user trust? Investigating trust in automation across different consumer-oriented domains in young adults, military, and older adults. *Theoretical Issues in Ergonomics Science*, 1–22. doi:10.1080/1463922X.2016.1175523

Which Anxiety Cues do Employment Interviewers Detect?

Deborah Powell

University of Guelph, CA

Contact: dpowell@uoguelph.ca

Employment interviews are one of the most popular methods for selecting employees (Huffcutt, Van Iddekinge, & Roth, 2011). Yet many people approach a job interview with palms sweating and heart racing, knowing they cannot put their best foot forward because of their anxiety. Anxious job candidates receive lower ratings of interview performance, and are less likely to be hired for the job (Ayres & Crosby, 1995; Feiler & Powell, 2013; McCarthy & Goffin, 2004). These lower scores have a human cost in terms of lowered self-esteem, and an organizational cost in terms of missed opportunities. Yet little attention has been devoted to understanding the factors driving the relation between anxiety and interview performance. We wanted to know *why* anxious interviewees receive lower ratings, and what behaviors (or cues) might signal anxiety to interviewers. Anxious interviewees may find it empowering to understand those behavioral expressions of their anxiety that lead interviewers to infer that they are anxious, versus those they should be less concerned about.

We investigated specific behavioral expressions (cues) of anxiety exhibited by anxious interviewees, which we called “micro cues” as well as trait judgments (e.g., honesty, warmth). The behavioral (micro) cues are precise and objective, consisting of counts (e.g., number of times touching one’s face) or duration (e.g., amount of time spent smiling) of specific behaviors. In contrast, trait judgments or macro cues (e.g., warmth, enthusiasm) are a product of engaging in several (discrete) behaviors simultaneously and take the context of the behavior into account. We examined these behavioral cues and trait judgments to identify the behaviors that make candidates appear anxious in a job interview.

The relation between anxiety-related behavioral cues and anxiety was best conceptualized with Brunswik’s lens model. The center of the lens model contains the “cues”. In this case, the cues are either broad trait judgments (e.g., attentive, dominant) or specific verbal and non-verbal behavioral cues (e.g., fidgeting, smiling). The left side of the model refers to the interviewees’ actual level of anxiety (as measured by self-ratings), whereas the right side represents observer judgments. The correlations between the interviewees’ actual level of anxiety and each of the cues is called *cue validity*, meaning there is a link between anxiety and a behavioral manifestation of that anxiety. The correlation between a cue and the observers’ judgment is called *cue utilization*; a correlation here implies that observers are using those cues to make their ratings. To the extent that there are valid cues, and the cues are utilized, there will be agreement between self and other ratings – called observer accuracy.

For our study, undergraduate cooperative education (“co-op”) students ($N = 125$) were recruited from a Canadian university. The interviewees completed a mock interview as part of their preparation for applying to co-op positions. Participants rated

their own level of interview anxiety immediately following their interview, and interviewers were asked to rate (observed) interview anxiety. The interviews were videotaped, and several sets of coders rated micro cues (e.g., smiling, fidgeting, touching face) and macro cues (e.g., attentive, confident, honest, likeable.)

Our goal was to investigate behavioral cues and traits that are exhibited by anxious interviewees, and those that are detected as anxiety by interviewers. Our results indicated that few behavioral cues significantly correlated with interviewee or interviewer ratings of anxiety. Four cues correlated with *interviewee* ratings of anxiety: hand gestures ($r = -.22$), nodding frequency ($r = -.20$), pause time ($r = .25$), and speech rate ($r = -.39$). Three cues correlated with interviewer ratings of observed anxiety: lick/bite lips ($r = .23$), speech rate ($r = -.36$), and torso movement ($r = .29$). Therefore, only one behavioral cue (slower speech rate) correlated with both interviewee and interviewer ratings of anxiety.

Unlike the behavioral cues, nearly all of the trait judgments significantly correlated with both interviewee and interviewer ratings of interview anxiety and also appeared to correlate with each other, suggesting there may be some commonalities. We conducted a factor analysis, and found two factors, which we called “Interpersonal Warmth” and “Assertiveness.” We found that the relation between *interviewee* and *interviewer* ratings of interview anxiety is mediated through Assertiveness (traits such as confident, dominant, optimistic).

Our findings suggest that low Assertiveness and slow speech rate are two key cues that are both indicative and revealing of interview anxiety. Our results also suggest that traits (vs. behavioral) cues offer more insight into what mechanisms are driving the interview anxiety-interview performance link. Therefore, anxious interviewees should try to extensively prepare for interviews in advance to lessen prolonged silences following questions and provide more details in their responses.

An important insight from our study is that interviewees may manifest their anxiety in different ways; there is no common pattern of cues present in all anxious interviewees. Often interviewees are worried that they are engaging in nervous tics, when in fact the impression that they convey of themselves as assertive (or not) appears to be more indicative of their anxiety. We hope that this research will serve as an important stepping stone to advance our understanding about the role of interview anxiety in the selection process.

Key reference:

Feiler, A. R., & Powell, D. M. (2016). Behavioral expression of job interview anxiety. *Journal of Business and Psychology, 31*, 155–171.

References:

Ayres, J., & Crosby, S. (1995). Two studies concerning the predictive validity of the personal report of communication apprehension in employment interviews. *Communication Research Reports, 12*(2), 145–151.

Feiler, A. R., & Powell, D. M. (2013). Interview anxiety across the sexes: Support for the sex-linked anxiety coping theory. *Personality and Individual Differences, 54*, 12–17.

Huffcutt, A. I., Van Iddekinge, C. H., & Roth, P. L. (2011). Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance. *Human Resource Management Review, 21*(4), 353–367. doi:10.1016/j.hrmmr.2011.05.003

McCarthy, J. M., & Goffin, R. D. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology, 57*(3), 607–637.

Unacquainted Callers can Predict which Citizens will Vote Over and Above Citizens' Stated Self-Predictions

Todd Rogers

*Harvard Kennedy School,
Harvard University, US*

Leanne ten Brinke

*Department of Psychology,
University of Denver, US*

Dana R. Carney

*Haas School of Business,
University of California, Berkeley, US*

Contact: todd_rogers@hks.harvard.edu

People are regularly asked to report on their likelihoods of carrying out consequential future behaviors, including complying with medical advice, completing their educations, and voting in upcoming elections. Although these stated self-predictions are notoriously unreliable, they are used to inform many strategic decisions. We report two studies examining stated self prediction about whether citizens will vote. We find that most self-predicted voters do not actually vote despite saying they will, and that campaign callers can discern which self-predicted voters will not actually vote. In Study 1 ($N = 4,463$), self-predicted voters rated by callers as “100% likely to vote” were two times more likely to actually vote than those rated unlikely to vote.

Study 2 ($N = 3,064$) replicated Study 1’s finding that untrained callers can discern which self-predicted voters will not actually vote. Additionally, Study 2 involved recording the interactions between citizens and the phone callers. This allowed us to use a Brunswikian lens approach to test which voice-related nonverbal cues callers relied upon to differentiate those who followed through from those who did not. The recordings were coded by three research assistants who were blind to voting behavior and hypotheses. Coders were trained to listen to each audio clip and code the presence/absence and qualitative aspects of nonverbal behaviors related to uncertainty, cognitive load, and arousal. Pairs of coders were randomly assigned to code each behavior. After practicing the to-be-coded behavior on a separate set of audio stimuli, both coders coded a randomly determined subset of 10% of the audio clips. After reliability was established, one coder proceeded to code all stimuli on that behavior. Behaviors were coded one at a time over 3 months. Table 1 lists all behaviors coded, definitions, references to relevant research, coding scales, approach, and inter-rater reliability. To determine how callers formed accurate predictions of which self-predicted voters would follow through or not, a Brunswikian lens model was fitted to the data (see Brunswik, 1956). Figure 1 reveals valid nonverbal behaviors (i.e., correctly used cues that led to accuracy), invalid behaviors (i.e., cues used that did not lead to accuracy), and missed opportunities (i.e., valid cues that were not leveraged). Sounding uncertain, sounding insecure, and having longer latencies prior to responding to the self-prediction question were valid

behaviors, meaning callers utilized these behaviors to make accurate judgments. In other words, these behaviors were correlated with both callers' predictions and citizens' actual voting behavior. Callers also interpreted sounding tense, and sounding nervous as signals that self-predicted voters would not vote, but these nonverbal behaviors were unrelated to actual voting behavior. Additionally, the more speech fillers self-predicted voters used the less likely they were to vote, though callers failed to use this diagnostic cue when predicting who would vote. Speech rate and mean and maximum vocal pitch were unrelated to both caller predictions and actual voting behavior.

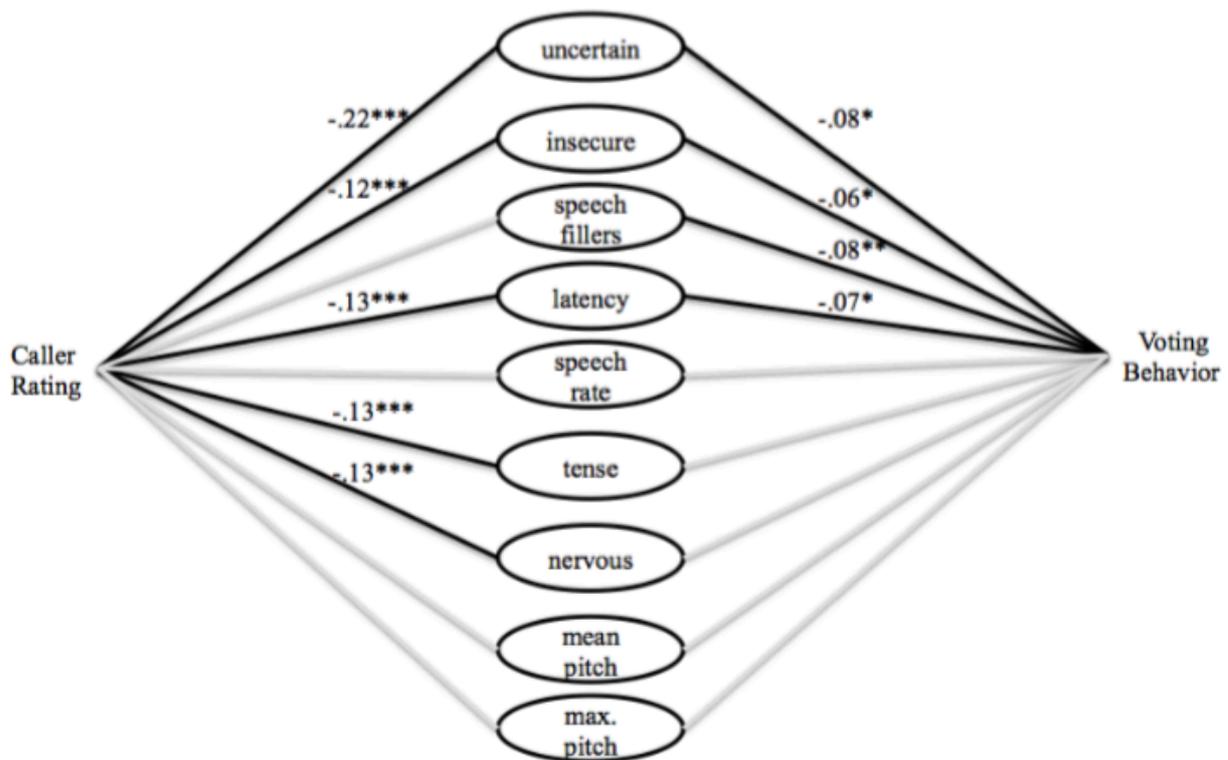


Figure 1. Brunswikian lens model for Study 2 (Rogers, ten Brinke, & Carney, 2016). Correlations between nonverbal cues and respondents' voting behavior (Right) and caller ratings of voting likelihood (Left). Black lines indicate significant relationships (*p < .05; **p < .01; ***p < .001); gray lines indicate nonsignificant relationships, *ps* > .05. Callers correctly incorporate in their vote predictions respondents sounding more uncertain, more insecure, and having longer speech latency/voice onset (these attributes were related to actual respondent voting). Callers incorrectly incorporate in their vote predictions respondents sounding more tense and more nervous (these attributes were unrelated to actual respondent voting). Callers correctly did not incorporate in their vote predictions respondents' mean vocal pitch, maximum vocal pitch, and higher speech rate (these attributes were unrelated to actual respondents voting). Callers missed the opportunity to incorporate in their vote predictions respondents' use of speech fillers (this attribute was related to actual respondent voting).

In conclusion, ordinary, untrained human judges can significantly improve predictions of who will follow through versus flake out on important commitments. This knowledge could increase the efficiency of the allocation of campaign resources and is likely to be valuable in other domains as well. For example, it could be used to improve the targeting of costly interventions that increase patient compliance in medical care – a context in which billions of dollars are wasted due to patients' lack of follow through – and to better identify the students most at risk of failing to follow through on their

college study and persistence plans (see Morgan, 2001). In short, the findings speak to a broad challenge in social life and suggest a simple input that leverages human social judgment to increase the accuracy of intervention targeting.

Table 1
Nonverbal Variables

Behavior	Definition	References	Scale	Approach	α
Nonverbal Variables Associated with More Uncertainty					
Uncertain	Subjective, global rating of unsure/uncertain/risky	27, 28	1 (<i>uncertain</i>) – 7 (<i>certain</i>)	Human	.95
Insecure	Subjective, global rating of feeling worried/unsure	27, 28	1 (<i>insecure</i>) – 7 (<i>secure</i>)	Human	.94
Nonverbal Variables Associated with More Cognitive Load					
Speech fillers	Utterances such as: “um”, “ah	23, 32	# (i.e., count)	Human + counter	1.0
Latency	Time before onset of respondent’s response	23, 32	Seconds	Human + stopwatch	.96
# of words	# of words uttered	23, 32	# (i.e., count)	Human + counter	.94
Duration	Length of time utterance lasts	23, 32	Seconds	Human + stopwatch	.84
Nonverbal Variables Associated with High Arousal					
Tense	Arousal through strict, staccato vocal qualities	23, 29	1 (<i>relaxed</i>) – 7 (<i>tense</i>)	Human	.71
Nervous	Arousal through shakiness and unease in voice	23, 29	1 (<i>comfortable</i>) – 7 (<i>nervous</i>)	Human	.77
Pitch	The mean and maximum vocal vibrations observed for an utterance	30, 31	Vocal frequency in Hz	Praat software	1.0

Note. The variable speech rate was calculated by dividing the number of words uttered (“# of words”) by the duration of the utterance, in seconds (“duration”; Hauch et al., 2015). Source: Rogers, ten Brinke & Carney (2016).

References:

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality & Social Psychology Review, 19*(4), 307–342.

Morgan, T. M. (2001). The economic impact of wasted prescription medication in an outpatient population of older adults. *Journal of Family Practice, 50*(9), 779–781.

Rogers, T., ten Brinke, L., & Carney, D. R. (2016). Unacquainted callers can predict which citizens will vote over and above citizens' stated self-predictions. *Proceedings of the National Academy of Science, 113*(23), 6449–6453. doi:10.1073/pnas.1525688113

What is Represented in a Vignette?

Lars Sjö Dahl

Lund University, Sweden

Contact: lars.e.sjodahl@gmail.com

Vignettes, often used as tasks in decision-making research, are occasionally discussed from a validity point of view: Does the vignette content represent the intended environment-subject relations, that is, how valid may these vignettes be? This problem could be approached in several ways, such as employing subject experts (Akhuly & Gupta, 2014, pp. 302–303). Built on Funder's realistic accuracy model presented in Hammond and Stewart (2001, p. 369), the following questions could be asked, all relevant for judging validity of vignette content.

- Are the mentioned behavioral cues relevant for the trait descriptions?
- Is it physically possible for the decision maker to observe the cues referred to?
- Has the decision maker been in a position to directly detect the cues in question?
- Do the cues referred to indicate covert behavior or behavior of episodic character?

All these considerations are illustrated in the vignette presented by Akhuly and Gupta (2014).

When discussing the internal validity of vignettes Hughes and Huby (2004, p. 36) ask the crucial question: "Have they been vetted by an expert panel?" Certainly this is an important request. However, in our Brunswikian context we will present an alternative approach, focusing on the representativeness of our vignette content. To illustrate, we started with defining an aspect or task-domain, in this case nurses' psycho-social role in contact with their patients (Sjö Dahl, 1974). This aspect domain was defined by 11 short descriptions of individuals' psychosocial needs, for example, the need for being accepted as a member of a group. The following three concrete examples were given: "Work together or be in contact with others; be able to maintain contact with those you are emotionally attached to; remain loyal towards friends, to experience affinity with each other" (Sjö Dahl, 1974, pp. 18–19).

An interview-guide was compiled and a large number of interviews with nurses on their places of work was carried out during a couple of years, all in accordance with Flanagan's (1954) critical incident method, briefly described by the following premises for its application.

No planning and no evaluation of specific behaviors are possible without a general statement of objectives.

Discussions have failed to emphasize the dominant role of the general aim in formulating a description of successful behavior or adjustment in a particular situation.

It is clearly impossible to report that a person has been either effective or ineffective in a particular activity by performing a specific act unless we know what he is expected to accomplish.

A basic condition necessary for any work on the formulation of a functional description of an activity is a fundamental orientation in terms of the general aims of the activity (Flanagan, 1954, p. 336).

The incidents, collected in our nursing research, are not short episodes confined to a few lines. On the contrary, they are usually extensive and full of contextual details (information) due to the complexity of the reported incident. This result has been achieved by prompting questions like the following:

- What you have told me now is very interesting and highlights many important points of view.
- How did you first become aware of this problem?
- Did you take any measures in this situation?
- Did you act in some way, later on?
- Could you think about other alternative actions, good or poor, to handle this situation?
- Why did you choose to act as you did?
- What was the consequence (the result) of the measures you took?
- Do you regard your own acting, handling of the situation, as satisfactory with regard to the patient's situation?
- Why do you think so?
- Did anyone else (except the patient) take part in this incident?

These interviews were comprehensive and contained strict as well as flexible, qualitative parts with open-ended questions.

The restrictions on the data-collection in a critical incident study confine the data, that is, behavioral context, to a predefined ecological field in terms of goals or purposes. This strategy may involve a tedious, preparatory workload sometimes involving cooperation between different subject experts. However this approach facilitates the ecological applicability of the achieved research results. A similar approach to constructing vignettes based on real cases can be found in the medical field (Skanér, Strender, & Bring, 1998; Skanér, Bring, Ullman, & Strender, 2000).

It should be admitted there are several reality conditions in our nursing study that are difficult to replicate in vignette studies, such as the following ones: 1) most nurses work under time pressure sometimes with both an emotional and a cognitive load. In our critical incident study we tried to respect that condition by always allowing the nurse to decide a suitable date, as well as time for and length of the interview: 2) in order to avoid a disproportional number of true, negative incidents the interviewer asked every second time for positive and negative incidents, respectively.

When applying Egon Brunswik's concepts (1952) to our discussion about the validity of vignettes one realizes that his conceptual world may be a bit difficult to translate into operational terms. This is somewhat remarkable as Brunswik's language mediates the impression of a writer trying to be as precise as possible. However, Brunswik's theoretical and conceptual world is not a fixed, rigid system. In fact, it lends itself to a variety of operational definitions and applications. To illustrate the meaning variations (scope of denotations and connotations) possible with some of Brunswik's theoretical concepts we take a closer look at the concepts 'proximal' and 'distal'. These concepts may refer to a wide range of operational meanings:

- the physical aspect (time and space). Time is an important variable in any decision situation and in validity testing;
- the mental, representational space-time dimension, freedom to move back and forth in our representational world;
- the concrete–abstract dimension, meaning that proximal cues or goals are more concrete than abstract ones, which are regarded as more distal;
- the dimension positive-negative valence, meaning that positive valence goes with proximal cues, negative valence goes with distal cues.

This need for meaning specification extends to both task descriptions and to the actual decision and judgment processes. It also includes criterion descriptions. So, for example, the following five questions are relevant not only to task descriptions, but also to the criterion specification.

- Is the task, the problem under consideration, timeless or is it of an occasional nature?
- Is the task of an applied or theoretical character?
- Is the task description complete or incomplete (more or less incomplete)?
- Is the goal inherent in the task unambiguous or is it open for various interpretations?
- How does the task relate to the universe it is supposed to represent?

Looking at the vignette world in relation to the real world, Evans et al. (2015, pp. 160–170) distinguish between the following three validity aspects. First, the vignette world is supposed to have a simulation function, intended to approximate a defined theoretical construct, like patient-centeredness, symptom-centeredness or human resources (Akhuly & Gupta, 2014). Second, the content of a vignette has an eliciting function: assumed to exist in the real world it elicits responses one would really make. This function relates to the vignette’s internal or content validity. A third request concerning vignette studies is that they should produce results that can be generalized to real-world situations. This generalizing function presupposes a description of a task domain to which the vignette results are intended to be generalized, a requirement in line with Egon Brunswik’s representative design concept.

The richness of Egon Brunswik’s conceptual world resembles an artist’s endeavor to avoid restricting the message to a predefined subject area, an ambition described by Hammond: “...that central states such as motivation, set, attitude, or personality characteristics should be varied, and that distal effects such as goal achievement should be observed. And here at last, we find those wide-arched dependencies which Brunswik found to be the ultimate source of problems for psychology...the dependencies between distal causes and distal effects” (Hammond, 1966, pp. 21–22).

References:

- Akhuly, A., & Gupta, M. (2014). Competency based recruitment decisions: A lens model approach. *The Psychologist-Manager Journal*, 17(4), 279–306.
- Brunswik, E. (1952). *The conceptual framework of psychology*. *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.
- Evans, S. C., Roberts, M. C., Keeley, J. W., et al. (2015). Vignette methodologies for studying clinicians’ decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology*, 15(2), 160–170.
- Flanagan, J. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358.
- Hammond, K. R. (Ed.) (1966). *The psychology of Egon Brunswik*. New York, Chicago, San Francisco, Toronto, London: Holt, Rinehart and Winston, Inc.

- Hammond, K. R., & Stewart, T. R. (2001). *The essential Brunswik. Beginnings, explications, applications*. Oxford, New York: Oxford University Press.
- Hughes, R., & Huby, M. (2004). The construction and interpretation of vignettes in social research. *Social Work & Social Sciences Review*, 11, 36–51.
- Sjödahl, L. (1974). *Att vårda människor. (Care of people)*. Stockholm: Natur och Kultur. (In Swedish).
- Skanér, Y., Bring, J., Ullman, B., & Strender, L.-E. (2000). The use of clinical information in diagnosing chronic heart failure: A comparison between general practitioners, cardiologists and students. *Journal of Clinical Epidemiology*, 53(11), 1081–1088.
- Skanér, Y., Strender, L.-E., & Bring, J. (1998). How do GPs use clinical information in their judgements of heart failure? A clinical judgement analysis study. *Scandinavian Journal of Primary Health Care*, 16(2), 95–100.

**Policy Capturing is More Resistant to Faking
than Other More Traditional Self-Report Techniques**

Alan J. Tomassetti, Reeshad S. Dalal & Seth A. Kaplan
George Mason University, US

Contact: atomasse@gmu.edu

The notion that policy capturing—and related techniques such as judgment analysis and conjoint analysis—may reduce a decision-maker’s ability to fake or to consciously respond in a socially desirable manner seems to be very common throughout social science research. However, this assertion was not adequately tested until our recent publication in *Organizational Research Management* (*Is Policy Capturing Really More Resistant Than Traditional Self-Report Techniques to Socially Desirable Responding?*). Here, we will summarize that work.

One reason researchers are drawn to lens model techniques is the ability to systematically control what information is being considered, and at times, it has been claimed that the indirect manner in which decision policies are assessed mitigates respondents’ ability to actively influence their responses in a manner that is more aligned with societal norms (e.g., Rynes, Schwab, & Heneman, 1983). However, the few previous studies that are cited as providing support for this claim (i.e., Arnold & Feldman, 1981; Brookhouse, Guion, & Doherty, 1986; Mazen, 1990; Rogelberg, Ployhart, Blazer, & Yonker, 1999) did not compare individuals’ responses when asked to answer honestly to their (within- and between-level) responses when asked to respond in a socially desirable manner.

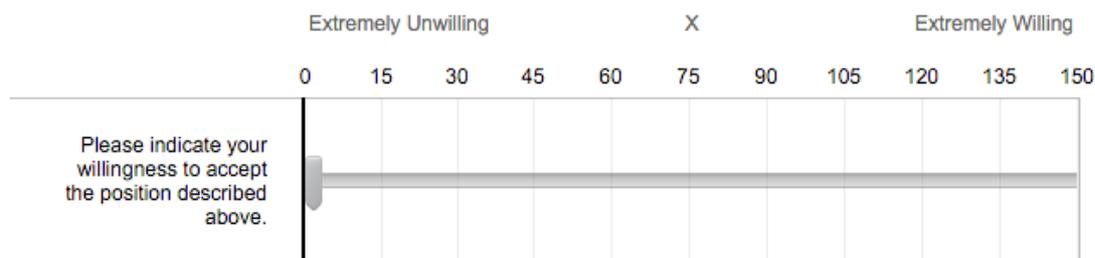
We sought to build from and expand this earlier work while directly assessing the degree to which policy-capturing studies actually mitigate decision-makers’ ability to fake their responses in a socially desirable direction. To do so, we used instructional sets in an attempt to directly manipulate decision-makers’ responses. Four separate instructional sets were used: no specific instructions (i.e., the control condition), instructions to respond in a socially desirable manner, instructions to respond honestly, and instructions warning that faking would be detected and consequences would result (i.e., withholding research credit; see Dwight & Donovan, 2003 for a meta-analysis on the effect of warnings). Indeed, comparing policy differences among and within decision-makers when instructed to respond in a socially desirable manner versus either (a) asked to respond honestly or (b) warned not to be dishonest provides the most stringent test of the policy-capturing’s ability to reduce faking. Furthermore, the comparison of policies under no specific instructions to the other conditions was also examined as a less conservative test.

The design chosen for this analysis asked respondents to indicate how willing or unwilling they would be to accept a given position described by six factors: type of work; opportunity to use important skills and abilities; amount of autonomy and independence; amount of responsibility and leadership; amount of salary and fringe benefits; and amount of flexibility in scheduling work hours and vacation (see Figure 1 for examples). Before testing the focal hypothesis (simply stated as “that policy-

capturing techniques are less susceptible than traditional self-report techniques to socially desirable responding,” p. 263), we conducted a pilot study to determine the rank ordered importance of our cues when participants are responding honestly as compared to when they are asked to respond in a socially desirable manner.

Please indicate your willingness to accept the job described below:

The organization provides essential services and products to the public. The job provides opportunities to use important skills and abilities. The job provides an above average amount of autonomy and independence. The job provides responsibility and leadership opportunities. The organization provides above average pay and fringe benefits. The organization provides flexibility in scheduling work hours and vacations.



Please indicate your willingness to accept the job described below:

The organization provides nonessential services and products to the public. The job does not provide opportunities to use important skills and abilities. The job provides a below average amount of autonomy or independence. The job does not provide responsibility or leadership opportunities. The organization provides below average pay and fringe benefits. The organization does not provide flexibility in scheduling work hours and vacations.

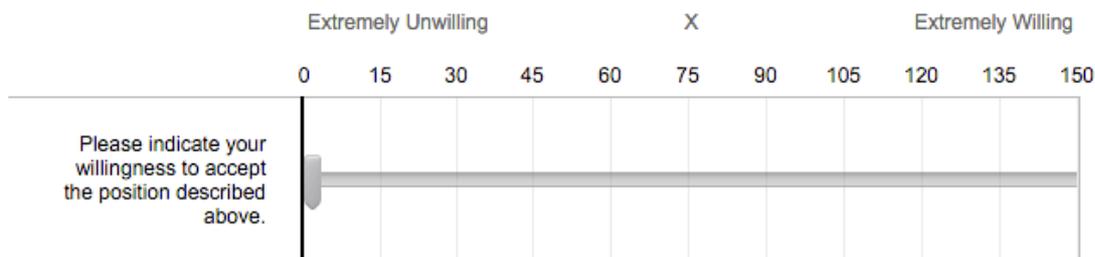


Figure 1. Two example policy-capturing profiles. Each profile consists of six cues, representing job and organization characteristics. One profile (top) contains all six cues with positive wording, whereas the other profile (bottom) contains all six cues with negative wording. Every profile in the focal study contained either the positive or the negative wording for each of the six cues, such that, across all the profiles, all possible cue wording combinations were used. Respondents were asked to read each profile and then indicate, via the slider, how willing they would be to accept the job described in that profile.

Once the separate policy patterns were determined, participants from a large public university were asked to participate in a two-wave study. Three hundred and thirty-six participants completed Wave 1, and one week later, 158 participants completed Wave 2 (47% retention rate). During each wave, participants responded to a policy-capturing measure and four more traditional self-report techniques (i.e., Likert-type, forced choice, ranking, and point-distribution). The policy-capturing fully crossed all six cues (64 unique scenarios) and repeated two studies in order to assess test-retest reliability ($r = .79$ and $r = .76$ for Waves 1 and 2, respectively; both $p < .001$). At each wave, participants responded to all five measures under the same instructional set condition. All participants received instructions to respond in a socially desirable

manner at one of the two time points and one of the other three instructional sets at the other time point.

Results were analyzed at both within- and between-person levels because within-subject designs are often seen as more transparent than between-subject designs (e.g., Tversky & Kahneman, 1983). For the sake of simplicity, the results will be presented in general terms here (see Tomassetti et al., 2016 for more detail). Differences between instructional sets within and across techniques were compared with omnibus tests followed by within-technique t-tests. These analyses assessed the degree to which decision makers were able to respond in a more socially desirable manner when instructed to versus when warned not to respond dishonestly or when instructed to respond honestly. Taken together, results indicated that across the four more traditional self-report techniques, the average responses under the socially desirable instructions were inflated in the socially desirable direction compared to average responses under the warning or honest conditions. This pattern of results, however, did not hold for the policy-capturing technique, thereby strongly—and empirically—supporting the notion that policy-capturing designs are less susceptible to faking than the more traditional techniques examined here.

In addition to providing clear support for the claim that policy-capturing is more resistant to faking than other methods, our study offers several avenues for future research that may be of interest to readers of this newsletter. Specifically, the mechanism by which policy-capturing studies resist faking is still unknown. As an auxiliary analysis, we examined the idea that complexity may be a factor, but our results only provided partial support for this idea. In addition, this study examined only one direction of faking (i.e., faking good) but neglected the other direction of faking bad. There is no clear reason that faking good and faking bad must be mirror opposites; in fact, research suggests that they may actually be distinct (Viswesvaran & Ones, 1999). As such, understanding how policy-capturing compares to faking bad offers a clear avenue of future research.

Finally, the authors would like to thank the Brunswick Society for their gracious invitation to contribute.

References:

- Arnold, H. J., & Feldman, D. C. (1981). Social desirability response bias in self-report choice situations. *Academy of Management Journal*, 24(2), 377–385.
- Brookhouse, K. J., Guion, R. M., & Doherty, M. E. (1986). Social desirability response bias as one source of the discrepancy between subjective weights and regression weights. *Organizational Behavior and Human Decision Processes*, 37, 316–328.
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16, 1–23.
- Mazen, A. M. (1990). The moderating role of social desirability, age, and experience in human judgment: Just how indirect is policy capturing? *Organizational Behavior and Human Decision Processes*, 45, 19–40.
- Rogelberg, S. G., Ployhart, R. E., Balzer, W. K., & Yonker, R. D. (1999). Using policy capturing to examine tipping decisions. *Journal of Applied Social Psychology*, 29(12), 2567–2590.
- Rynes, S. L., Schwab, D. P., & Heneman, H. G. (1983). The role of pay and market pay variability in job application decisions. *Organizational Behavior and Human Performance*, 31(3), 353–364.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293–315.
- Tomassetti, A. J., Dalal, R. S., & Kaplan, S. A. (2016). Is policy capturing really more resistant than traditional self-report techniques to socially desirable responding? *Organizational Research Methods*, 19(2), 255–285.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197–210.

To get information about future Newsletters and the Brunswik Society you are welcome to join (free of charge) the email list of the Brunswik Society at:

<http://www.brunswik.org/>