THE RELIABILITY OF THREAT ASSESSMENT JUDGMENTS IN EXPERTS, HUMAN

RESOURCE PROFESSIONALS, AND NOVICES: A RELIABILITY STUDY

by

TIMOTHY DAVID BRIGHAM

(Under the direction of Robert Mahan)

ABSTRACT

Policy capturing of threat assessment judgments was performed on "Experts" (including law

enforcement and military professionals), Human Resources (HR) professionals, and a student

sample. Policies were analyzed idiographically using the Brunswikian Lens Model, followed by

policy aggregation for nomothetic comparisons of factors such as cognitive consistency in

execution of judgment policies and utilization of particular cues. Hypotheses included that ability

to control, execute, and describe policies would be greatest in Experts, and that concealment and

race cues would be used more by experts. It was also hypothesized that experts would differ least

from each other in cue utilization. Mixed results were found. Surprisingly, many hypotheses

regarding performance were met by the Human Resource professional Group, but not the Expert

group. Discussion of possible reasons and suggestions for future research follow.

INDEX WORDS: Threat assessment, Threat evaluation, Security assessment, The Lens
                 Model, Brunswik, Brunswikian, Multiple Cue Probability Learning, Social
                 Judgment Theory, Law enforcement, military, Human Resources, Threat,
                 cognitive consistency, cognitive control

THE RELIABILITY OF THREAT ASSESSMENT JUDGMENTS IN EXPERTS, HUMAN

RESOURCE PROFESSIONALS, AND NOVICES: A RELIABILITY STUDY

by

TIMOTHY DAVID BRIGHAM

B.A., University of West Florida, 2000

M.S., The University of Georgia, 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

THE RELIABILITY OF THREAT ASSESSMENT JUDGMENTS IN EXPERTS, HUMAN

RESOURCE PROFESSIONALS, AND NOVICES: A RELIABILITY STUDY

by

TIMOTHY DAVID BRIGHAM

Major Professor:	Robert Mahan
Committee:	Adam Goodie
	Keith Campbell

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2009

DEDICATION

Dedicated to my Family for their unwavering support throughout my life and education; As well as to Adrienne Gallo, Peter Schaefer and Robert Mahan, without whom this document would not exist.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Assessment of security threats has taken on a new degree of importance following the events

of September, 11, 2001. On February 2, 2009 a 'worldwide alert' from the U.S. State

Department advised heightened caution for Americans to exhibit vigilance in assessing threats

from 'extremists,' and advising a 'heightened state of awareness.' Although the alert goes on to

cite previous attacks using Improvised Explosive Devices (IEDs) and piracy, little practical

advice, besides heightened vigilance, is given on how situational awareness of threat probability

can be practically improved. The assessment and quantification of threat in the U.S. has

developed from diffusely defined initiatives primarily aimed at developing countermeasures

against mass casualty terrorism. This failure to adequately classify and delineate threat properties

has led to the absence of prescriptive protocols. Yet it is clear that relevant and timely threat

assessment doctrine must be useful, scalable, and well delineated. The effects of such initiatives,

the practicality of assessment, and the effects of training and experience are among the issues

examined in the current study.

Threat assessment holds great promise for determining vulnerability and guiding

interventions in potentially lethal and high impact situations. A common goal of all threat

assessment programs is to prevent an attack on a potential target. However, the art of combining

threat related data and information in the form of an assessment; particularly in real time, almost

always requires and element of personal common sense on the part of the assessor (McPhaul,

London, Murrett, Flannery, Rosen, & Lipscomb, 2008). Results of the effects of experience and

training have been understudied, and it remains unclear how training affects threat assessment. It is important to begin specifying how formal training affects the process of real time threat assessment? Further, how might the procedures used by experts differ from those organizations that opt for less rigorous approaches to counter-violence policies? Is threat assessment a general process that is similar in different contexts and between people facing dissimilar threats? Law Enforcement Officers (LEO) and Military Police (MPs) are trained at assessing security threats and are considered 'Experts' for our purposes by that definition, but does that training refine a process that exists generally in other groups, such as Human Resource (HR) professionals who must make safety decisions regarding hiring and managing personnel, and in laypeople? Experts, HR professionals, and laypeople also likely differ on how often they must make such assessments. In the case of LEOs and MPs, constant vigilance and judgment after judgment regarding a potential threat may be required for an entire 8 hour work shift. Human Resources professionals, alternatively, less often are required to make assessments regarding the security risk or threat posed by a potential employee.

Experts and Expertise

Many of the questions that have recently arisen about training and threat assessment derive from the literature of differences between experts and novices. Bransford, Brown and Cocking (1999) identified what they believe to be key principles defining experts' performance and knowledge. They include experts' abilities to notice meaningful patterns that go unnoticed by novices, deep and organized content knowledge, and that the knowledge is often applied in an almost procedural manner rather than a step by step analytical one. Glaser (1992) argued that real-world experience is a necessary component of expertise. Ericsson and Charness (1994) argue that alongside real world

experience, study, practice, self-monitoring and intentional self-improvement are among the most important aspects in the making of an expert.

Experts seem to have a much deeper understanding of their subject matter, and that knowledge is often applied in a quick, even automatic manner (Jacobson & Wilensky, 2006). While experience and practice leads to formation of a policy or response, real world experience seems to be the key to becoming a true expert (Bransford, Brown and Cocking 1999; McPherson, 1999).

The Mixed Message Regarding Profiling

Criteria for threat and security judgments often differ between organizations. For example, some experts (specifically Police and Transportation Safety Screeners) are explicitly told to ignore race when evaluating threat, and that is the official policy for many law enforcement organizations (Anonymous U.S. Policeman, personal communication, July, 2008; Transportation Safety Authority Website, 2008). Others have suggested openly that *when properly used*, profiling can and should be utilized as a tool; viewed as good detective work, good behavioral analysis and good statistical sense. For example, profiling was suggested in the social science literature by Secret Service officers to identify potentially or likely violent students (Reddy, Borum, Berglund, Vossekuil, Fein, & Modzeleski, 2001). Profiling is also advocated in the HR realm as a way to minimize threat from employees (for example see *Profiling the Lethal Employee;* Kelleher, 1997). Though debate continues over the appropriateness and usefulness of profiling in general, and certain information (such as race) as predictors in particular, whether such information is made use of and by whom, is one of the issues addressed in the current study.

Modeling Multi-Cue Threat Assessment: A Brunswikian approach

A modeling principle applicable to the manner in which multi-dimensional threat assessment can be understood is based on the Brunswikian judgment protocol. The approach has been developed

within the context of the Brunswikian notion of probabilistic functionalism and vicarious mediation in

visual perception (Brunswik, 1956). Brunswik's view was that visual perception is the activity

characterized by a perceiver interacting with his/her ecological environment; an environment whose

tendency it is to distribute or "scatter its effects". Within this context, he viewed the important

ecological dimensions (or cues) of the environment as being probabilistic and not fully reliable or

dependable. The fact that the environment presents the perceiver with redundant information in the

form of correlated cues (i.e., the environment is vicariously mediated), means that the perceiver must

wisely select and use the cues most diagnostic of a given behavioral or perceptual goal. A rather good

functional example of the meaning of the probabilistic nature of environmental cues is taken from

Gordon (1989) on Brunswikian Psychology:

> "Suppose we are searching for an edible fruit. Let us assume that edible fruit is
>
> (a) darker, (b) redder, (c) softer and (d) sweeter. Obviously, darker and redder are
>
> visual cues, softer is tactile, sweeter is gustatory: the environment is scattering its
>
> effects. And these cues, the only ones available, are all imperfect: all carry some
>
> risk. Not all ripe fruit is red, nor is all red fruit edible. Sweetness often indicates
>
> edibility, but some poisonous fruits are sweet. Some fruit is less edible when soft,
>
> some fruit will be rotten." (pp. 131)

As a functionalist, Brunswik's basic perceptual theme was adaptive in nature. That is, in order to

survive the perceiver must deal in risk and uncertainty by acting like an intuitive statistician. Thus, the

perceiver must be able to (a) select meaningful cues from a plethora of ecological information, (b)

factor the riskiness of the situation, (c) combine the cues and risk factors, and (d) render a judgment

leading to action (e.g., avoid the thicket of trees else risk being eaten by a tiger).

Brunswik (1952, 1956, 1957) was responsible for introducing a formal systems approach to the study of human cognition. From his probabilistic perspective on the relationship between a perceiver and his/her environment came Brunswik's lens model of behavior, which defined the structural characteristics of the person/ecology relationship. This unique model, with both normative and descriptive features, defines the complex multidimensional representation of human behavior within an ecological context (Brunswik, 1952, 1956).

Social Judgment Theory

The lens model has since been since modified and expanded in order to represent a general model of human judgment and decision-making (Brehmer and Joyce, 1988; Hammond and Adelman, 1976; Hammond, McClelland, and Mumpower, 1981; Tucker, 1964); while the regression ideas fist conceived by Brunswik and Hammond have found formal shape (Hoffman, 1960). These developments, refinements, and application of Brunswikian, and specifically lens model, methodology, 20 years after the initial lens model was conceived, have resulted in what has been termed Social Judgment Theory (SJT). To be accurate, SJT is not a theory; it makes no predictions. Rather it is a metatheory, providing a framework for examining and testing judgments in relation to probabilistic cues. It is also primarily an idiographic statistical approach that focuses on information from an individual judge making judgments based on numerous cue arrays. In this way it is atypical in that it is not itself a nomothetic analysis tool or metatheory, yet if a researcher so chooses, it can provide data that can be analyzed in a nomothetic manner.

Policy Capturing in The Lens Model

Figure 1 illustrates the policy capturing mechanisms in the regression-based lens model analysis. The lens model essentially distinguishes between an object or condition that is defined by various information sources (cues), and the psychological representation of the object or condition which is

defined through a particular judgment policy. The lens model portrays the environment as a series of cues whose relationships with the environment are less than perfect. A decision maker is viewed as interacting with his or her environment through a "lens" which is often distorted because of this imperfect and uncertain relationship. The relationship between the cues and the environment is typically characterized by "ecological validities" that, in theory, can range in absolute value from 0 to 1.0. Ecological validity represents the predictive importance of each cue. The manner in which a decision maker uses particular cues can be modeled by a regression equation that predicts an individual's judgment of an object from a linear combination of cue weights. The degree to which a decision maker accurately assesses the characteristics of an object or condition in the environment is expressed by the correlation between the object's true values and those predicted by the decision maker (Hammond and Wascoe, 1980).

The lens model provides a formal means for quantifying the influence of various task features on human cognitive behavior. As can be seen in Figure 1, the lens model provides the means for manipulating various properties of the task, and additionally provides a network of task and cognate behavioral descriptive terms that can be useful in locating a person's cognitive activity on the continuum. Figure 2 presents a form of the lens model used when a criterion is unknown.

The lens model represents the modeling approach taken to evaluate and assess judgment reliability and competency, as well as forming aggregation indices that highlight organizational judgment behavior and effectiveness.

Research Focus and Hypotheses

The Lens Model equation guided the manner in which we characterized the experimental hypotheses. Because of the absence of a known criterion, the judgment analysis utilized the single system Lens Model (Figure 2). Discussion of the method and logic of the cue selection

process appears in chapter two, while cues and their potential values are presented in Appendix A. Cues selected for use were Age, Gender, Apparent Race, Verbal Information ('Accent'), Mannerisms, Concealment Potential ('Concealment'), and Dress Style.

H1. Since LEOs and Military ('Experts') had been trained on assessing threat, and on average were more experienced than the other groups (HR Professionals and 'Students' approximating laypeople; discussed and defined in Chapter 3) thru training and actual engagements with threat events, it was hypothesized that Experts would demonstrate higher cognitive consistency *(*r_S in the lens model) or an ability to more reliably execute their judgment policies than the other groups.

H2: Experts have better insight thru their training and experience into how they explicitly use their policy information over HR Professionals and Students. Therefore, Expert judges should show smaller differences between their subjective and functional weights (captured policies) than do HR Professionals and Students.

H3: Expert law enforcement officers are trained to ignore race to avoid profiling charges which potentially risk U.S. civil rights violations (e.g., Gabbidon, Marzette, Peterson, 2007*;* Garrett, 2001). However there is some evidence to support that profiling exists in LEO populations due to the saliency of ethnic individual difference values. Therefore it is hypothesized Experts will systematically weight the Race cue in their captured threat policies.

H4: Further, when asked to acknowledge the use of Race, they will discount the cue's relevance in their subjective (self described) policies.

H5: As a consequence of programmatic training of Experts on threat assessment, Expert participants should demonstrate higher homogeneity on narrative threat assessment than the

other experimental groups. Therefore Experts' captured policies should be more similar to the group average policy than the similarity to average policy in the other groups.

H6: From a law enforcement perspective, the potential for concealment of weapons is particularly important in overall threat assessment activity. Therefore Experts and HR Professionals should place a higher weight on the Concealment cue than Student participants.

TRUE WORLD STATE   JUDGED WORLD STATE
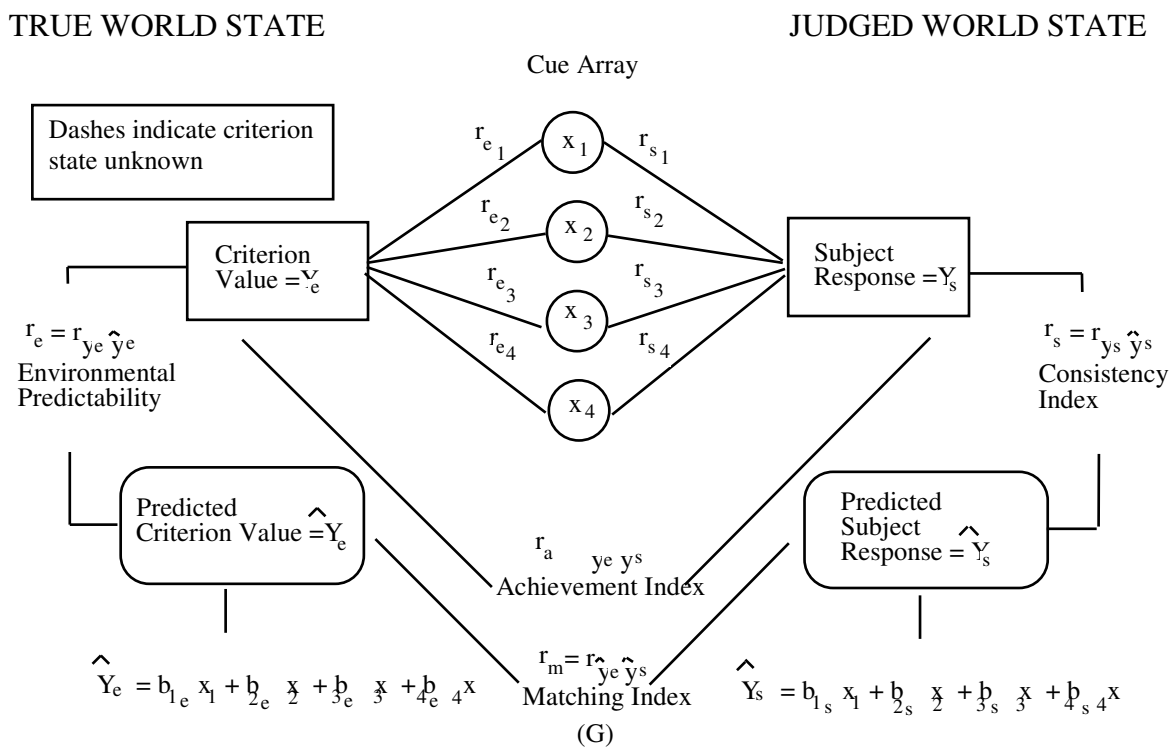

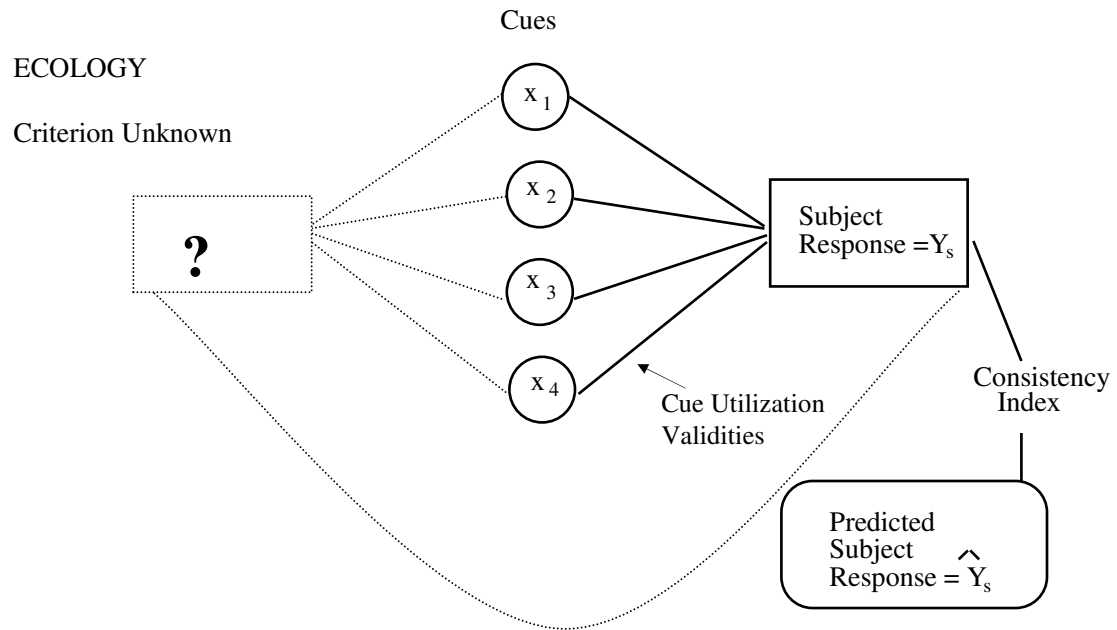Figure 1. Double System Lens Model

Figure 2. Single System Lens Model

CHAPTER 2

JUDGMENT POLICY CAPTURING

Policy capturing was performed on participants representing three samples: Law Enforcement Officers, Homeland Security Personnel, and former front line military personnel and Military Police made up our Experts, HR Professional (defined as employees charged with involvement in the hiring or firing process), and Students (from a large university psychology subject pool). Policies were first examined idiographically in order to assess individual reliability indices and relative frequency of significant use of particular cues. We also averaged polices and lens model indices and made comparisons between groups when possible. Because of the absence of a true criterion, the judgment analysis focused on the 'Judged World' side of the lens model in Figure 1, represented alone in Figure 2.

In policy capturing, it is required that idiographic analysis, in which data from individual participants is analyzed, be performed before any group (nomothetic) analysis can take place (Cooksey, 1996). Each participant's judgments on each trial are important in creating an individual participant's judgment profile, which includes the participant's R^2, beta weights indicating their usage of the cues and their associated significance levels, and the overall significance of the model for that participant. Following such idiographic analysis, beta weights and R^2 values can be aggregated, an average regression equation can be created that represents the group, and nomothetic analyses can be performed.

Relevant Predictors

Ecological considerations are a key aspect of Brunswikian philosophy and methods. This is particularly demonstrated in Brunswik's (1957) notion of representative design where he indicts traditional experimental methods with failing generalization symmetry. Here he notes that experimental statistical methods focus on random sampling in order to support participant population inferences, however, in most cases these methods fail at insuring inferences concerning predictor variable populations and representations of ecological parameters. This failure impacts context generalizations, and according to Brunswik, is often in part responsible for the absence of cumulative scientific findings and unified theoretical frameworks.

The cues were selected on the basis of citation frequency in the homeland security literature of threat evaluation. Random sampling predictor variables (cues) is often difficult to achieve in practice (see Mahan and Marino, 1996, for an example), being highly mindful of ecological characteristics is a hallmark component to Brunswikian methodology.

Having sampled from the homeland security enforcement literature for our cues, it is possible to find that cues utilized by these Experts are not the cues utilized in making threat assessments by HR Personnel and Students. Such a finding might be interpreted as evidence that threat assessment in a law enforcement or national security context is different qualitatively from threat assessment in other contexts, including everyday life, either because the cues used by Experts are not naturally used by, or are used differently used by, the untrained. A quantitative difference in cue use may suggest that formal training (and not just experience as with the HR Professional group) is the most important factor in creating a threat assessment Expert.

Paper People Narratives as the Vehicle of Cue Presentation

Judgment Analysis studies have typically presented simple lists of cue values to subjects and asked them to make some judgment based on them (Brehmer & Joyce, 1988). More recently the use of what have been termed 'paper people', in which trials are presented in the form of a descriptive narrative format containing a set of cue values, has been gaining popularity. Narratives have been argued to be more ecologically valid in most cases than cue value lists (Woehr & Lance, 1991). They impose less interpretational demands than lists of cues, and approximate reality more than numeric cue value lists. Narratives were thus chosen as our method of stimuli presentation.

The creation of each trial, or each paper person, is performed by selecting a cue value for each cue from the previously created list of cue values to be utilized. Use of Microsoft Excel allows for easy creation of randomized combinations of cue profiles. For example, the cue 'Age' would have all its values listed in a column or sheet from which one value is randomly chosen. The process continues until a value for each cue is randomly chosen and a trial or paper person has been created. The individual values are then meshed together (either automatically or by hand) into narrative form (see Appendix B).

Controlling for cue intercorrelations so that cues are (at most) lowly correlated is also possible. Each level of each cue can be crossed with unique levels of the other cues, so that each a cue value does not co-occur more frequently with any particular other cue value.

CHAPTER 3

METHODS

Participants

The lens model based paradigm utilized in the current study is inherently idiographic and allows significant amounts of data to be obtained from a single subject in the form of criterion judgments based on predictor cue profiles. The ideographic analysis uses individual judgments as the unit of analysis. Aggregation (nomothetic analyses) can be performed only after thorough analysis of individuals in policy capturing studies such as the proposed (Cooksey, 1996).

The 3 conceptual groups from which we sampled, security Experts, Human Resource Professionals, and Students, were recruited using a variety of methods. The University of Georgia Psychology Department maintains a subject pool from which the Student population was obtained. HR Professionals were sought out via personal contacts and word of mouth referral, as well as requesting participation from appropriate internet message boards and mailing lists. The Expert group was obtained by word of mouth referrals and posts/requests to appropriate internet message boards and mailing lists.

A total of 80 usable Student participants, 20 HR Professionals and 20 Experts participated. Students were asked to indicate their race, though it was possible to leave this blank and continue the experiment. Other demographics were not recorded, in part for anonymity reasons. However, some reasonable speculations about the Student sample based on known characteristics of large southeastern university psychology program populations (Colleges and Universities in Kentucky, 2009). Females tend to compose 2/3rds of the population. Whites or Caucasians typically make up

82% of the Psychology student population, African-Americans 4%, Asian or Pacific Islanders 6%, Hispanic/Latino 2%, and 5% are 'other' or 'multi-racial'. "Traditional" college students, which make up the mass of the UGA Psychology student population, range in age from 18-23, and a majority are native to the state in which the university is located (UGA Profile, 2008).

Similarly, factors regarding the Expert and HR Professional groups can be estimated based on researcher interaction with some participants, and characteristics typically associated with these groups (i.e. Experts, requiring training and experience to reach that role, could be estimated to be at least roughly 10 years older than Students, on average). Meeting members of the Expert group helped support the estimate that they were more typically older than students, and more often male. The HR Professionals was more difficult to make estimates about, particularly because researchers had contact with only 4 of the HR Professional participants. However, all were female (as have been the majority of HR Professionals the researcher has had personal experience previously), and were closer in age to the Expert group than the Student group. It was also known that several of the HR Professionals came from a large consulting firm.

Finally, participants who were self-described Experts were asked to briefly indicate their training or experience. There was a nearly equal distribution between military or homeland security personnel, and domestic police who had received training in terrorism analysis and response (i.e. 40% and 60%, respectively).

Materials

Seventy-five 'Paper People' narratives were created. Web based data collection (via SurveyMonkey) was utilized to conduct the study in order to both ensure participant anonymity and increase ease of participation. Because of the absence of a true criterion, cue values for the generated quasi-randomly while controlling for intercorrelation as discussed above. This method

producing low cue intercorrelations helped facilitate interpretation of regression parameter estimates.

Procedure

Potential participants were directed to a website which contained University of Georgia institutional review board informed consent information for them to print and retain, as well as email and phone contact information to direct questions. Because the only record of a participant's involvement in the study would have been a signed informed consent form (i.p. addresses were hidden by the web based data collection service per our request), the most reasonable compromise between anonymity and safety appeared to be to require participants to read an online informed consent and indicate they have done so by typing a keyword ('accept') in a box on the webpage before they are able to continue on to the study. Because of the need for students to retain proof of participation for credit hour purposes, they were asked to print and sign the final page of the study, which served as a receipt that was provided to the researchers for the purpose of awarding credit to student participants.

Instructions then appeared on the screen and an example paper person and sample judgment was provided. This page would not advance until the participant indicated they understood the paradigm. Participants were asked to make careful judgments, but not to dwell too long on any one trial.

Following the instructions and example, paper people narratives were presented one at a time, each with a box in which the participant is asked to indicate their judgment on the threat criterion ranging from 0 (no threat) to 100 (certain threat) until 75 had been performed. Following all paper person trials, participants were then asked to distribute 100 points amongst a list of the

cues, indicating their diagnostic importance to the participant's judgment process. Finally,

participants were asked to indicate their occupation.

CHAPTER 4

RESULTS

Reliability Measures.

The lens model analysis was used as the framework for generating the dependent measure used to assess performance. R^2, or (as explained below) r^2_s, is the <u>cognitive consistency index</u>, which represents the ability of the participants to control the execution of their knowledge regarding the judgment task. Consistency can be best conceptualized as the difference between participant judgments and the predicted judgments generated from their multi-cue regression policies, signified by r^2_s. Calculated differently but yielding the same result, R in the regression model is the multiple correlation between cues and judgments, and is mathematically the same index of how reliably judges executed their judgment policy. Perfect consistency would mean that no difference exists between predicted judgments from the captured regression policy and actual judgments. The squared version of R is typically used to simplify discussion of variance accounted for, and how reliable a participant is, in percentage form

Regression parameters were standardized to clearly demonstrate the diagnostic utility of participant cue use.

The subjective weights were defined by the participants at the conclusion of their participation by being instructed to "indicate how important each piece of information was in making your decisions by distributed 100 points between the factors listed below" (Anderson, 1981; Cook & Stewart, 1975; Cooksey, 1996).

A Bootstrap Variant Applied to the Student Sample

Sample size discrepancies forced the use of a statistical bootstrapping methodology for nomothetic comparisons (see Hartigan, 1975 for review). Gibbons, Olkin and Sobel (1999) have argued bootstrapping is appropriate in situations in which unequal sample sizes are not due to the effects of a treatment variable. We opted for this method for comparison simplification purposes, and because we had no specific hypotheses regarding individual beta weights within the student sample. Four equal samples of 20 were taken from the student group. Their R^2 means and variances were examined for representativeness of the larger sample, and then averaged to compose a single set of 20 student cases (see Ghosh, Vogt & Vogt, 1988, for a discussion of this method). We compared the mean R^2 and variance to the original larger group (of 80) and their similarity is presented in Table 1.

Idiographic-Descriptive Analysis

Table 2 presents idiographic data, including relative frequencies of significant R^2s, and relative frequencies of significant individual cues. Results of particular interest are discussed here. Experts exhibited significant R^2s in 90% of cases. This was followed closely by HR professionals with 85%, and students with 80%. Interestingly regarding the idiographic analysis, Race was significant in 15% of the Experts, however in those cases the standardized betas associated were relatively high: .358, .411 and .554. Race was further significant in 35% of HR professionals, and 5% of students. The Mannerisms cue was a significant standardized beta in 85% of experts, 70% of students, and 80% of HR Professionals. And the potential for concealing an object cue ('Concealment') was a significant standardized beta in 60% of experts, 35% of students and 45% of HR Professionals.

Nomothetic-Inferential Analysis

Table 3 presents group average values for standardized betas, R^2 values, and variances. The Mannerisms cue had the largest effect on judgment for members of all three groups (see also figure 3). Interestingly, Concealment was the 2nd most significant factor for experts, whereas Gender was the next significant factor in students and race was 2nd in the HR Professionals (with Gender running a very close third).

H1: To test our first hypothesis that Experts would demonstrate a greater degree of cognitive consistency or ability to reliably execute their policies over HR Professionals and Students, an F test of group average R^2 was performed. No statistically significant differences were found, $F(2, 57) = 1.719$, $p=.189$.

H2: To test the hypothesis that Experts would indicate their subjective policies with the most accuracy to their captured policies, i.e. exhibit the most insight, compared to other groups, the following steps were performed: 1. In each group, each participant's subjective cue weight, for each of the seven cues individually, was subtracted from their captured cue, yielding 7 difference scores per participant. 2. Each difference between captured and subjective cue weight for each individual and each of the 7 cues was squared. 3. Each participant's squared difference scores were summed. 4. These squared, summed, difference scores were compared between groups using an ANOVA. However, no significant differences were found, $F(2, 57) = .457$, $p=.636$.

H3 and H4: An omnibus F test on the captured weights for the Race cue indicated a significant difference between groups, $F(2, 57)=6.7$, $p=.002$. However Tamhane's T2 post hoc test indicated the omnibus result to be driven by a difference between Student and HR Professionals' use of the Race cue, $p=.01$, and did not involve the Expert group, as we had hypothesized. Because of H4's prediction that Experts would subjectively weigh Race as

unimportant, which would differ significantly from their captured policy weights, was contingent on the result of H3, we did not proceed to attempt to test H4.

H5: To test the hypothesis that Experts would exhibit the most homogeneity in policies, the following process was followed. 1. For Each individual in each group, the individual's captured cue weights were individually subtracted from the respective group average cue weight, for all seven individual cues. 2. These differences between individual cue weights and group average cue weights were squared. 3. The squared differences were summed for each participant. These differences were tested to determine if there was less within group difference from the average policy for the Expert group. An ANOVA indicated a significant difference between the groups in homogeneity of policies, $F(2,57)=7.1$, $p=.001$. Tamhane's poct hoc test indicated the differences to be between Students and HR ($p=.012$) and HR Professionals and Experts ($p=.047$)   Means of the sum of squared calculated differences from average policy indicated Students (M=.17, S.D.=.034) and Experts (M=.174, S.D. =.102) were similar, while HR Professionals (M=.0885, S.D.= .093) were the standouts with the lowest difference from group average policy.

H6: The hypothesis that the 'Concealment' cue would be utilized more by the Expert and HR Professionals than the Students was tested by an ANOVA of captured policy Concealment cue weights. A significant difference was detected, $F(2, 57)= 8.388$, $p=.001$. Tamhane's post hoc indicated a difference between LEO and both HR and students $p=.049$ and .03 respectively, supporting our hypothesis.

Table 1. Descriptive Statistics of Student Bootstrap, Subsamples and Original Sample

Original Sample of 80 Student Participants		
Mean R^2	.53	
R^2 Variance	.031	
Subsamples of 20 Participants per group		
Subsample 1		**Subsample 3**
Mean R^2 .478		Mean R^2 .498
R^2 Variance .039		R^2 Variance .035
Subsample 2		**Subsample 4**
Mean R^2 .544		Mean R^2 .598
R^2 Variance .027		R^2 Variance .024
Bootstrap of 20 Cases		
Mean R^2	.53	
R^2 Variance	.031	

Table 2. Idiographic Data

	Relative Frequency of Significant R²	Relative Frequency by Group of Significant Cue Standardized Betas						
		Age	Gender	Race	Accent	Mannerism	Concealment	Dress Style
Experts	90%	10%	25%	15%	5%	85%	60%	5%
Students	80%	15%	45%	5%	5%	70%	35%	10%
HR	85%	20%	30%	35%	5%	80%	45%	5%

Table 3. Average R² and Standardized Betas per group

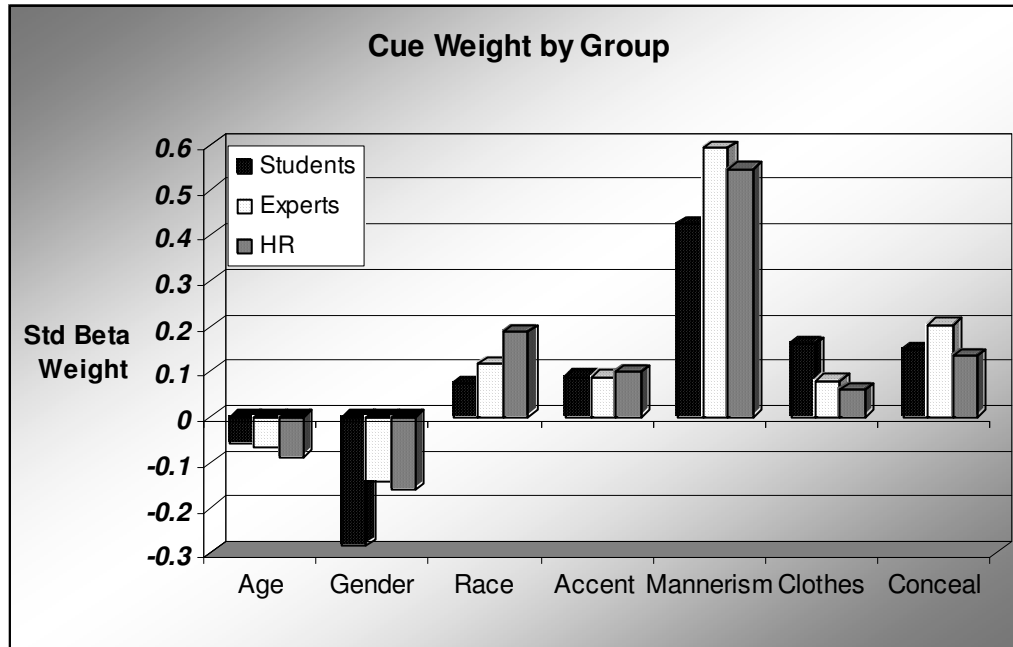|	Students	Experts	HR
R²	.525*	.557*	.592*
R² Variance	.03	.021	.0098
Age	-.057	-.066	-.088
Gender	-.284*	-.141	-.162
Race	.074	.118	.187*
Accent	.089	.088	.100
Mannerisms	.425*	.593*	.546*
Dress	.162	.079	.058
Concealment	.148	.203*	.133

* indicates p<.05

Figure 3. Cue Weights by Group

CHAPTER 5

DISCUSSION

Discussion of some of this study's interesting findings and supported hypotheses, as well as examination of potential reasons for failure to find support for some of the proposed hypotheses (regarding cognitive consistency, expectation of group policy similarity being highest in the Expert group, and some surprising results from the HR Professional group), are now examined. It was hypothesized that cognitive consistency would be greater in the Expert group due to training and experience, but group differences were not found. The finding could be interpreted as training and experience having a null effect on consistency of policy application in Experts. Alternatively, given that the narratives were merely an attempt to simulate a real potential for a threat assessment, Experts may have felt that being consistent in applying their policies was of less than usual importance, or key cues may not have had their usual impact. They furthermore might typically use a policy involving very different cues, and thus in the study changed their policy depending on what was most salient in a given scenario.

In regards to hypotheses about specific cue usage, the results were mixed. The finding that the Mannerisms cue was the cue most utilized by all groups indicates nervous or suspicious behavior was thus often a useful cue in making a threat assessment. Though this is somewhat intuitive, nervous behavior can occur for many reasons, and may not necessarily indicate a threat (someone being shaken by personal news they just received for example). If a criterion for threat were available, it would also be possible to find that visible nervousness is at times not an

indicator of someone who proceeds to commit an act of violence (hijackers buying tickets or mass murderers waiting for the opportunity to begin their carnage, for example).

The finding that Experts made more use of the Concealment cue supported hypothesis six, and would seem to make sense given the particular nature of the Expert group's work. This would appear to be the case regardless of whether the participants are domestic police officers, or military personnel overseas. Potential for concealing an object provides obvious safety concerns for Law Enforcement, as well as military personnel.

No support was found for the hypothesized differences between groups in ability to correctly state their own policy (i.e. demonstrate policy insight). The lack of support for this hypothesis would on the surface indicate students are as good as security professionals at 'telling what they know.' A possible explanation for these findings is that requests for subjective cue weights was perceived to be purely a test of 'profiling' and that respondents were thus not always honest. In support of this idea, a member of the Expert group emailed feedback stating the interpretation that "the study appears to be about profiling." Although profiling was an issue of interest as discussed above, this was not the main thrust of the study, but the perception that it was our intended focus may have affected Experts' performance. A second, potentially more important alternative explanation is that, as the reviewed expertise literature discussed (i.e. Bransford, Brown and Cocking 1999), Experts may be applying policies in a procedural manner, which is difficult to translate into declarative knowledge. Discussion of how this might related to future research follows in the final section.

Related to the above finding, the relatively high difference between Students and other participants' use of the Race cue may have resulted from Experts' fear of appearing to profile. Students made the least use of this cue in comparison to other groups. A potential explanation is

that variability in race is higher in the Student sample than the other groups, and that more interactions with students of different races lead to this cue being seen as less useful. Age differences and ideas potentially associated with age differences might likely be at work as well.

In regards to the Gender cue use findings, that the Gender cue (based both on beta weight and significant cue ratios by group) appeared more diagnostic for Students and HR Professionals (whose cue weight approached significance), the results were not to the degree expected based on what is known about gender ratios in violent workplace offenders or college campus sexual offenders. However, the pattern of results, while not in each case reaching statistical significance, did reflect the pattern of expected use.

Participant gender might be relevant as well. All Expert group members who had researcher contact were male, whereas the few members of the HR Professionals who had direct researcher contact were female. As discussed below, future research should include a more thorough and complete collection of demographic data.

Heterogeneity in Experts?

A possible confound that may have affected the findings of the current study was the heterogeneity of the members of the Expert group. Some described themselves in the optional 'comments' section of the survey as having military experience in security, while others were police officers with training in security and the assessment of threat.

Reflecting upon initial interviews with security and threat assessment experts during preparation for this project, there were very different opinions expressed by police officers and a former U.S. serviceman in Iraq and former Military Policeman (though at the time it was attributed to the serviceman being a personal friend of the researcher. and thus, more open, which also in part lead to the hypotheses regarding subjective and actual use of the Race cue). It

was also particularly suggestive that idiographic level analysis indicated the highest individual standardized betas of the 20 Expert subjects were both military (.511 and .410). Future research done in this area should consider the issue of homogeneity regarding the operational definition of 'Experts.' It would be especially useful to include a variety of classes of Experts in separate groups. These might include Military Police, front line serviceman based on region, police officers further subdivided by rank or duty, and so on.

On the Performance of Human Resources Participants

The performance of the HR Professionals, particularly regarding their individual similarity to the group average policy was of great interest. Perhaps the hiring/firing and employee supervision processes leads to more practice and a 'fine tuned' policy. It may be the case that HR Professionals do more evaluations of threat than might be estimated. However, another possibly important alternative explanation was brought up by an HR Professional in contact with the researchers. HR Professionals may not necessarily make more threat assessments, but the ones that they do make often involve hiring or firing decisions based on highly personal (often unpleasant or even threatening) interactions which may lead to learning which is qualitatively different from, for example, security personnel who are often observing a variety of people at once or are focused on the security of a place or object. Such a difference may eventually be found to involve differences in memory encoding, potentially based on high emotionality of the event. Often in such scenarios, an incident involving an employee or potential employee might also circulate via word of mouth and become the focus of attention for a period of time, whereas an interaction of the same nature might be considered typical for military or law enforcement.

Suggestions for Future Research

Future research should attempt to address the limitation of written narratives. Though not a limitation specific to this study, using videos or photos might prove more ecologically valid and lead to support for hypotheses similar to those in the current study. They may capture, at least to a greater degree, the essence of what a participant experiences during an actual assessment of threat. Relatedly, cues might be created so that there are both orthogonal (low or non-correlated) and inter-correlated cues, which could be analyzed separately, because a limitation of randomized or orthogonal cues is that they sometimes result in cases which seem unrealistic to judges (Cooksey, 1996).

The results of this study's insight measurements, taken into account with the literature on procedural-like behavior by some experts, and the emotional content apparent in some HR events, leads to suggestion that future work might benefit from a further examination of insight (or the lack thereof) in certain groups or individuals. Continued measures of insight and, to the degree possible, an examination of the process of policy execution (perhaps including response time) seems worthy of future attention.

The choice of cues utilized might also be varied in similar future research. Feedback indicated some potentially useful cues that could replace the ones which appeared to be of little use by participants. Researchers may include possible explanatory cue values for the Concealability cue values (such as congruence of clothing with weather, or carrying papers in a briefcase) which might mitigate or change use of that cue, for example. Further, one Expert suggested that the destination and origin of individual paper people would have been informative; for example is the person traveling from a known drug source and arriving at a

known distribution point? This may, however, complicate attempts to apply the same cue profiles and narratives to different groups.

Since no criterion measure was available, and thus there were no matching or achievement indices to investigate, it might also be interesting and reasonable to consider analyzing data such as that resulting from this study utilizing the policy agreement form of the Lens Model discussed by Cooksey (1996). In such a scenario, Experts might comprise left side of the double sided lens (which is usually the ecological side) and comparisons with other groups and their similarity in policies and cue use could be investigated by placing them on the right side. Collection of more demographic data might also be used to perform more thorough analyses on individual participants, as well as selected sub-groups based on factors such as gender or race.

Ideally, a follow-up to the current study would again utilize a layperson-like control group such as students, real world professionals who may face threats (such as human resources professionals), and a variety of sub-classifications of 'experts.' More demographic info could be useful in idiographic and nomothetic analyses, and more realistic stimuli might elicit more realistic responses, and help answer some of the questions which went unanswered in this study.

REFERENCES

Anderson, J.A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society, Series B (Methodological), 46(*1), 1-30.

Bransford, J.D., Brown, A.L, and Cocking, R. (eds.) (1999). *How People Learn: Brain, mind, experience, and school*. Washington: National academy press.

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

Brunswik, E. (1941). Perceptual size-constancy in life situations. *Psychological Bulletin, 38*, 611-612.

Brunswik, E., & Herma, H. (1951). Probability learning of perceptual cues in the establishment of a weight illusion. *Journal of Experimental Psychology*, 41, 281-290.

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in functional psychology. *Psychological Review, 62*, 193-217.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley: University of California Press.

Brunswik, E. (1957). Scope and aspects of the cognitive problem. In H. Gruber, R. Jessor, & K. Hammond (Eds.), *Cognition and the Colorado symposium*, (pp. 5-31). Cambridge: Harvard University Press.

Cook R.L., & Stewart, T.R. (1975). Comparison of seven methods for obtaining subjective descriptions of judgmental policy. *Organizational behavior and human performance, 13*, 31-45.

Cooksey, R. W. (1996). The methodology of social judgment theory. *Thinking and Reasoning, 2*(2), 141 - 174.

Ericsson, A.K., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist. 49*, 725–747.

Hartigan, J.A. (1975). *Clustering algorithms*. New York: John Wiley & Sons, Inc.

Jacobson M.J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences, 15*(1), 11–34.

Kelleher, M.D. (1997). *Profiling the lethal employee*. Santa Barbara: Greenwood Publishing Group.

Kentucky: Top five largest colleges and universities by student population (2008). Retrieved April 10, 2009, from Kentucky Colleges Universities and Technical Schools. http://education-portal.com/articles/ Kentucky_(KY):_Colleges,_Universities _and_ Technical_Schools.html

Mahan, R., and Marino, C. (1996). Delayed feedback decision making. *Human Factors Proceeding, 38,* 566-571.

McPherson S.L. (1999). Expert-novice differences in performance skills and problem representations of youth and adults during tennis competition. *Journal of Sport & Exercise Psychology*,*70*(3), 233-51.

McPhaul, K.M., London, M., Murrett, K., Flannery, K., Rosen, J., and Lipscomb, J. (2008). Environmental evaluation for workplace violence in healthcare and social services. *Journal of Safety Research, 39*(2), 237-250.

Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57,* 116-131.

Reddy, M., Borum, R., Berglund, J., Vossekuil, B., Fein, R., & Modzeleski, W. (2001). *Psychology in the Schools, 38*(2), 157-172.

Stewart, T. R. (1990). A decomposition of the correlation coefficient and its use in analyzing forecasting skill. *Weather and Forecasting, 5*(4), 661-666.

Transportation Safety Authority Website, (2008). *The 'why' behind air security*. Retrieved October 1, 2008. http://www.tsa.gov/press/happenings/usa-today-chertoff-qa.shtm

Tucker, L. R (1964). A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. *Psychological Review, 71*, 528-530**.**

UGA profile (2008). Retrieved April 9, 2009 from The University of Georgia web site http://www.career.uga.edu/employers/ugaprofile.html.

Woehr, D. J. & Lance, C. E. (1991). Paper people versus direct observation: An empirical examination of laboratory methodologies. *Journal of Organizational Behavior,12*(5), 387-397.

APPENDIX A

CUE VALUES

- Age
    - late teens-early twenties
    - twenties-thirties
    - forties-fifties
    - over fifty-five
- Gender
    - male
    - female
- Race/ethnicity appears to be
    - Caucasian
    - African-American/black
    - Asian
    - Arabic
    - Mixed race/undetermined
    - Latino
- Verbal information
    - You notice a heavy accent that is not American or European
    - They do not appear to speak English
    - They speak English but with a heavy American or European accent
    - They speak English with a Heavy Asian Accent
    - They peak English with a Heavy Arabic Accent
    - They speak grammatically perfect English with little accent
- Mannerisms
    - Normal
    - Slightly nervous
    - Very nervous
- Clothing style
    - Western/American/European
    - Asian
    - Arabic
- Concealment
    - Their clothes are well fitted.
    - Their clothing is baggy/bulky.
    - They are carrying a bag.

APPENDIX B

SAMPLE PAPER PERSON

The person you observe appears to be a male in his late forties or early fifties. He is conversing in English with a thick accent of unknown origin. His appearance suggests he is Latin American. He is wearing typical Western/European style clothing. His mannerisms seem slightly nervous. He is carrying a case.