ATTITUDES TOWARD HOLISTIC AND MECHANICAL JUDGMENT IN EMPLOYEE SELECTION: ROLE OF ERROR RATE AND FALSE POSITIVE AND FALSE NEGATIVE ERROR

Maya Yankelevich

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2010

Committee:
Scott Highhouse, Advisor
Amelia S. Carr
Graduate Faculty Representative
Milton D. Hakel
J. Devin McAuley
ABSTRACT

Scott Highhouse, Advisor

The overwhelming evidence in the literature favors mechanical judgment over holistic when making selection predictions. To date, no research has examined how the risk of error and the type of error in employee selection may impact attitudes toward holistic and mechanical judgment. Applying the principles of Error Management Theory (Haselton & Nettle, 2006), the goal of this research was to understand how the likelihood of specific types of selection errors (false positives versus false negatives) and the risk of these errors influence people’s attitudes toward holistic and mechanical judgment. Error rate (10% versus 40%) and type of error were investigated experimentally. A sample of 323 working adults took part in an experiment where they assumed the role of head of Human Resources for a large organization. Results of a fully crossed between-subjects design indicated an effect of error rate, but no effect of type of error on Perceived Usefulness of the selection procedure. There were also no interaction effects of judgment approach (holistic versus mechanical) and error rate or type of error. With the exploratory variable Perceived Legality, there was no effect of error rate, but there was an effect of type of error. The selection procedure was perceived to be less legal when false negative error was emphasized, as opposed to when false positive error was emphasized. Implications of these findings and future directions are discussed.
To my family and friends. Thank you for your love, support, and encouragement.
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Scott Highhouse, for his guidance throughout the dissertation process and during my time at BGSU. I could not have asked for a better advisor. In addition, I would like to thank my dissertation committee; Drs. Milton Hakel, Devin McAuley, and Amelia Carr for your thought-provoking contributions to my dissertation. Your willingness to share your perspectives was invaluable. I must also thank my friends, without whom my time in Bowling Green would not have been as inspiring; May Colatat, Anna Zarubin, Alison Broadfoot, Gabriel De La Rosa, Ashley Guidroz, and Wendy Fogo. Thanks!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>CONTENT</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Evidence of the Preference for Holistic Judgment</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Understanding the Preference for Holistic Judgment</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Role of Error in Resistance to Mechanical Judgment</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Understanding Reactions to Error</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Defining error</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Error Management Theory</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Varying costs of error</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Cost of error in selection</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Present Study</td>
<td>21</td>
</tr>
<tr>
<td>II</td>
<td>METHOD</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Participants</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Stimuli and Procedure</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Measures</td>
<td>24</td>
</tr>
<tr>
<td>III</td>
<td>RESULTS</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Preliminary Analyses</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Hypothesis Testing</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Exploratory Analyses</td>
<td>28</td>
</tr>
<tr>
<td>IV</td>
<td>DISCUSSION</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Exploratory Analyses: Perceived Legality</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Limitations</td>
<td>35</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES/TABLES

<table>
<thead>
<tr>
<th>Figure/Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Description of Sample</td>
<td>61</td>
</tr>
<tr>
<td>2</td>
<td>Results of Principal Axis Factoring on Dependent Variables</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>Means, Standard Deviations, and Intercorrelations for Study Variables</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>Summary of Analysis of Covariance: Effects of Judgment Approach, Error Type, and Error Rate on Perceived Usefulness</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>Marginal and Cell Means and Standard Deviations by Independent Variable for Perceived Usefulness, with Age as a Covariate</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>Summary of Analysis of Covariance: Effects of Judgment Approach, Error Type, and Error Rate on Perceived Legality</td>
<td>66</td>
</tr>
<tr>
<td>7</td>
<td>Marginal and Cell Means and Standard Deviations by Independent Variable for Perceived Legality, with Education as a Covariate</td>
<td>67</td>
</tr>
<tr>
<td>8</td>
<td>Non-significant interaction of judgment approach and error rate</td>
<td>68</td>
</tr>
<tr>
<td>9</td>
<td>Non-significant interaction of judgment approach and error type</td>
<td>69</td>
</tr>
</tbody>
</table>
CHAPTER I: INTRODUCTION

The study of error is not only in the highest degree prophylactic, but it serves as a stimulating introduction to the study of truth.

―Lippmann (1922, p. 409)

Rejecting evidence-based practice, human resource practitioners and hiring managers have consistently demonstrated a preference for unstructured interviews—and more generally, selection techniques that are subjective, intuitive, and based on personal evaluation (Dipboye, 1997; Lievens, Highhouse, De Corte, 2005; Terpstra & Rozell, 1997; Van der Zee, Bakker, & Bakker, 2002). Extant research, on the other hand, has established that traditional interviews are less predictive of successful job performance than selection decision aids, such as structured interviews, tests, and work samples (Schmidt & Hunter, 1998). Consequently, Highhouse (2008) lamented that possibly the greatest failure of Industrial-Organizational psychologists has been the inability to convince employers to use decision aids in hiring.

Managers’ preference for intuitive, over analytical judgment may foster the resistance to mechanical decision-making and evidence-based practices (see Highhouse, 2008). Shrivastava and Mitroff (1984), for example, suggested that managers prefer more subjective and experiential data to objective, measurable, and verifiable data when making organizational decisions. In fact, a 2002 survey conducted by executive search firm Christian and Timbers, currently CTPartners, revealed that nearly half (45%) of the corporate executives in their sample relied more on instinct than data to run organizations (Bonabeau, 2003).

Over half a century after Meehl’s (1954) seminal book, *Clinical Versus Statistical Prediction*, which detailed the superiority of mechanical judgment in predictive decision making, people still continue to prefer holistic methods. Numerous explanations have been offered for
why people resist using mechanical judgment (Dawes, 1979; Highhouse, 2008; Meehl, 1986; Sieck & Arkes, 2005). These explanations can be both motivational in nature, such as overconfidence (Sieck & Arkes, 2005), or they can be cognitive, such as belief in prediction expertise (Highhouse, 2008). The proposed study focuses on a cognitive factor largely overlooked in past selection research: beliefs about and reactions to error. In building the foundation for this research, I first discuss the evidence base suggesting that people favor holistic judgment. I then consider a number of explanations offered for the preference for holistic judgment, and especially the notion that beliefs and reactions to error influence judgment preferences. I apply the principles of Error Management Theory (Haselton & Nettle, 2006) to understand how selection errors (false positives versus false negatives) and the risk of these errors influence people’s attitudes toward holistic and mechanical judgment.

**Evidence of the Preference for Holistic Judgment**

An extensive research base spanning over 70 years has demonstrated that mechanical judgment methods are almost always superior to holistic (Grove, Zald, Lebow, Snitz, & Nelson, 2000), although people still seem to prefer holistic judgment (Corey & Merenstein, 1987; Hilton & Simmons, 2001; Lievens et al., 2005; Terpstra, 1996). One example of the disregard for mechanical judgment is the resistance of clinicians and tribunal counsels—clinical decision-making committees—in using actuarial aids in decisions to release mental patients. Hilton and Simmons (2001) examined the final release decisions of 189 tribunal counsels and found that even when empirically validated and highly effective risk assessment reports were available (in 119 of 189 hearings), counsels largely ignored the reports when making release decisions, preferring clinicians’ judgments instead. In fact, whereas the study found that the correlation between clinicians’ testimony and tribunal decisions was quite high ($r = .84$), the relation with
actuarial risk assessment scores and tribunal decisions was quite low ($r = .06$). Furthermore, even when clinicians were provided with the actuarial risk assessment scores, such information did not have an impact on their overall judgments of imminent risk and recidivism. It is unfortunate that this aid is not used given that if release decisions were based solely on the actuarial decision aid, only 24% of patients would be expected to recidivate, contrasted with the expected recidivism rate of 48% when the aid is ignored (Hilton & Simmons, 2001).

Research in the medical arena also suggests that people resist using decision aids. In the first part of their research, Corey and Merenstein (1987) developed a highly accurate decision aid to distinguish patients with ischemic heart disease from those without it. The relation between the aid’s predicted probability of cardiac ischemia and observed probability was $r = .93$, a high level of precision. Immediately following the portion of their research in which use of the decision aid was mandatory, the researchers made the aid available at the discretion of physicians. During this second period, the aid was used on only 2.8% of the patients, even though the second part of the study was conducted with the same physicians and in the same emergency room where the initial research occurred. Similarly, Graham et al. (2001) reported that only 31% of physicians in the United States report using the Ottawa Ankle Rule, a simple but highly effective decision rule for the prediction of ankle injuries. Focusing on patient reactions to mechanical judgment, Arkes, Shaffer, and Medow (2007) found over a series of studies that there is widespread patient resistance to diagnostic decision aids, which may drive their lack of use by physicians. In a study of auditors, Ashton (1991) found that only 2.2% consistently used a decision aid in a bond rating task.

Across a variety of selection situations, aversion to decision aids and mechanical judgment is also strong. Terpstra (1996) sampled 201 HR executives about the perceived
effectiveness of nine selection methods in predicting future job performance. His results indicated that the executives perceived interviews to be more effective than any paper-and-pencil assessment procedure (e.g., cognitive ability tests, personality tests, and biographical information blanks). The executives rated paper-and-pencil methods to be mediocre or below average in effectiveness, compared to above average effectiveness ratings for the interview. The marked preference for intuitively-derived candidate information is counter to established research showing that structured procedures, such as psychometrically sound general cognitive ability tests, are among the best predictors of job performance (Schmidt & Hunter, 1998; Schmidt, 2002). Despite beliefs to the contrary, cognitive ability tests, work samples, integrity tests, job knowledge tests, and structured interviews are all more predictive of job performance than unstructured interviews. Moreover, research has found that people cannot accurately assess personality characteristics predictive of job success with an unstructured interview (Barrick, Patton, & Haugland, 2000). Measures of conscientiousness, however, predict above and beyond cognitive ability to a greater degree than unstructured interviews—18% increase in validity when conscientiousness tests are included compared to an 8% increase when unstructured interviews are included (Schmidt & Hunter, 1998). Furthermore, meta-analytic results established that any one of the commonly used paper-and-pencil tests (i.e., cognitive ability tests, personality tests) alone outperformed the unstructured interview for predicting performance in entry-level and sales jobs (Huffcutt & Arthur, 1994; Vinchur, Schippmann, Switzer, & Roth, 1998). To the extent that an interview is structured, however, it may actually provide incremental prediction above cognitive ability and personality (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000), but it is this structure that people seem to resist.
Experimental evidence has also revealed a preference for unstructured selection procedures among organizational practitioners and job applicants. Using policy capturing, for example, Lievens, et al. (2005) assessed the value retail store managers placed on different kinds of applicant information. They presented participants with candidate profiles containing information about cognitive ability and personality, alternating how the information was gathered—interview or paper-and-pencil test. The results indicated that managers placed more weight on cognitive ability than on personality information when cognitive ability was derived using an interview than when it was derived using a paper-and-pencil test. Similarly, managers placed more weight on personality than cognitive ability when personality was derived using an interview than when it was derived using a paper-and-pencil test. Their results suggested that managers prefer candidate information predictive of successful performance to be gathered with intuitive methods, versus more structured mechanical methods. Brooks, Guidroz, and Chakrabarti (2009) investigated applicant reactions to different methods for incorporating diversity in employee selection decisions. They found that people in an applicant role, similar to managerial decision makers, also prefer more holistic judgment methods.

Diab, Pui, Yankelevich, and Highhouse (2009) recently presented findings of an experimental study of over 400 working adults that detailed people’s perceptions of tests and mechanical methods of data integration. Diab et al. found that participants perceived interviews as more useful, comprehensive, professional, congenial, and legal than tests assessing the same characteristics. Likewise, participants perceived the holistic method for combining information as more useful, comprehensive, professional, congenial, and legal than the mechanical method.

The above research suggests that people generally prefer unaided holistic judgment and perceive it more favorably than mechanical judgment. To better understand resistance to
mechanical judgment and the preference for holistic judgment we must also understand why the preference exits. In the following sections I discuss a number of explanations offered to account for the preference for holistic judgment.

**Understanding the Preference for Holistic Judgment**

Numerous researchers have speculated on the reasons for why people resist using decision aids in a variety of different settings (e.g., Dawes, 1979; Highhouse, 2008; Meehl, 1986; Yates, Veinott, & Patalano, 2003). Meehl (1986) listed seven psychosocial contributors to the continued reliance on holistic judgment among mental health professionals. These factors apply equally well to other fields where the choice between holistic and mechanical judgment can be made (e.g. medicine, selection).

The first factor Meehl (1986) listed was *ignorance*. He contended that many people are not aware of the research and data, do not understand the relevant mathematical or statistics information, and may not be aware of the fact that a controversy between these approaches even exists. Similar speculation has occurred as a result of the general neglect to implement evidence-based human resource practices (Gannon, 1983; Rynes, Colbert, & Brown, 2002). Offering a viable reason for the *ignorance* factor, Rynes et al. (2002) speculated that the lack of awareness and understanding might be the result of the complexity of information in research journals, making it too esoteric for many practitioners.

The second factor Meehl identified is *threat of technological unemployment*. As he explained, people may fear that their more complex holistic approach will be replaced by simpler technology (i.e. simple linear formula). Third is the idea of *self-concept*. This factor identifies the holistic judgment approach as an integral part of the professional’s identity. For example, auditors may feel it is their professional duty to be able to make bond ratings, just as physicians
may feel it is their professional duty to make diagnoses, without the help of decision aids.

Fourth, Meehl indicated *theoretical identification* as a factor. He compared people who identify with the holistic approach to those who identify themselves as Freudians: although they acknowledge that Freudian theory does not help in predicting anything of practical importance about patients, they nevertheless continue using it just because they are “Freudians.” The fifth factor Meehl noted is the *dehumanizing flavor*. This factor explains that using a formula to predict human behavior is equating a person to a lab rat, or an “it” as opposed to a “who.” The sixth factor is the *mistaken conception of ethics*. Similar in concept to the previous factor, Meehl explained that with the mistaken conception of ethics, what feels right to people takes precedent over greater predictive accuracy.

Finally, Meehl explained that *computer phobia* might be a factor in motivating holistic judgment preferences. This is characterized by people’s resistance to the idea that a computer can do things (i.e. predict outcomes) better than the human mind. Dawes’ (1979) idea of cognitive conceit seems compatible with this final factor offered by Meehl. Davis and Kottermann (1994, p. 59) defined Dawes’ proposition of cognitive conceit as “the illusion that the environment is more predictable than it really is and that greater cognitive effort will lead to better predictions than those afforded linear models known to be predictive.” Dawes was one of the first to suggest that people may resist mechanical judgment because of the error they perceive to be caused by that approach (Dawes, 1979). Other researchers have also offered explanations, which at their core seem compatible with Meehl’s computer phobia or Dawes’ cognitive conceit. One such explanation is termed the overconfidence hypothesis.

The underlying assumption of the overconfidence hypothesis is that people have confidence in their intuitive judgment, as compared to the performance of mechanical judgment
aids, and this overconfidence contributes to decision aid neglect. Whitecotton (1996) found support for this hypothesis in her study of professional financial analysts. Her results revealed a significant negative relation between a priori confidence ratings and reliance on a decision aid.

Arkes, Dawes, and Christensen (1986) provided further support for the hypothesized role of overconfidence in people’s resistance to decision aids. In their study, participants were instructed to select which baseball player, out of three, won the Most Valuable Player award (MVP) in 20 different seasons. The authors found that participants who scored better on a previously administered baseball quiz relied less on the provided decision aid (with a stated accuracy rate of 70%) in the subsequent task, and consequently, performed worse. The more knowledgeable participants were also more confident in their performance on the task. These results suggest that the more knowledgeable participants’ overconfidence reduced their reliance on the decision rule and thus weakened the quality of their decisions.

Improving on the limitations of previous correlational work, Sieck and Arkes (2007) found direct experimental evidence for the overconfidence hypothesis. In a series of studies, they provided participants with calibration feedback during a prediction task and found that the feedback reduced overconfidence and increased reliance on the decision aid, which led to improved judgment performance. The implication of these studies and others like them is that overconfidence is a real psychological phenomenon, possibly driving resistance to mechanical judgment (See also Koehler, Brenner, & Griffin, 2002; and Juslin, Winman, & Olsson, 2000 for opposing view).

Whereas overconfidence is a motivational factor, Highhouse (2008) offered two possible cognitive explanations for the resistance to mechanical judgment. He proposed that people hold two implicit beliefs that inhibit adoption of selection decision aids: the belief that (1) prediction
of behavior can improve with experience, and that (2) it is possible to reach near perfect prediction of performance. The myth of expertise, as Highhouse (2008) described it, is based on the assumption that people can become skilled at making intuitive judgments about others. Therefore, experts’ unaided judgments appear to be more trustworthy than judgments reached using decision aids. This idea resonates with the *self-concept* and *mistaken conception of ethics* explanations offered by Meehl (1986). Not only should “experts” be able to make judgments on their own, but it may also seem appropriate for them to do so.

Arkes et al.’s (2007) study with physicians provides additional support for Highhouse’s (2008) proposition that belief in expertise drives resistance to mechanical judgment. The researchers found that physicians who used a mechanical judgment approach during diagnosis were perceived as less competent than physicians who did not. Furthermore, the physicians using an aid were rated as having lower diagnostic ability and professionalism, and participants whose physicians used an aid were significantly less satisfied with their overall examination experience. Diab et al.’s (2009) finding that participants perceive expert judgment as more useful, comprehensive, professional, congenial, and legal gives further credence to Arkes et al.’s results.

Discussing his time with the National Science Foundation (NSF), Arkes (2008) recalled one biologist’s reaction to the proposition that the NSF use mechanical judgment in judging the merit of grant proposals. The biologist was shocked that “impersonal numbers” could be superior to the holistic judgment of an expert. The rigid formula, he asserted, could not take into account the various nuances that the expert could. The complexity of this resistance phenomenon is quite evident in this anecdotal example where the strong belief in expertise is compounded by the resistance to the idea that a computer can do better than the human mind (Meehl, 1986).
The second belief Highhouse (2008) discussed concerns people’s perceptions about prediction. Highhouse suggested that people believe if an applicant is right for the job, and is accurately and completely assessed, then successful performance is guaranteed. This is characteristic of Einhorn’s (1986) description of the clinical approach to prediction, which assumes that perfect predictability is possible in theory. However, not all failures in prediction are a result of the assessment process. Einhorn (1986) offered one explanation for this misapprehension. He distinguished between people with a deterministic worldview, who reject that the future is inherently probabilistic, and people with an analytical worldview, who accept uncertainty (and error) as inevitable. According to Einhorn, the holistic judgment approach is reflective of a deterministic worldview, which does not allow for “random error” and, as a result, sees causality in all associations. The development of superstitions, for example, may be rooted in the deterministic worldview, or the inability of people to assign causal probabilities to the events occurring around them (Foster & Kokko; 2009; Shermer, 1998). The inability to accept random error results in causal associations being associated with non-causal ones. Therefore, a determinist will argue that inaccurate prediction is a result of not having all the pertinent information at the time of prediction, not random error.

The key underlying assumption of the cognitive approaches detailed above seems to be the certainty people have in how accurate holistic judgment is in predicting outcomes. To the extent that people believe that accurate prediction is the result of thorough and complete knowledge of all details and that holistic approaches are the only way of achieving this complete knowledge, there will be a continued preference for holistic judgment. Unfortunately, the consequences of inaccurate prediction (i.e. cost of error) are often overlooked when a judgment
approach is selected. How people react to error and error costs, as a function of using holistic or mechanical judgment in a selection process, is the focus of this research.

**Role of Error in Resistance to Mechanical Judgment**

Although not extensively investigated, there is some support for the idea that knowledge of the error rate—or predictive accuracy—is related to resistance to mechanical judgment. Powell (1991) empirically demonstrated that people differentially reacted to mechanical aids as a result of their predictive (in)accuracy. Powell asked 216 participants to perform the same baseball task used by Arkes et al. (1986). In a between subjects experiment, Powell found that the more a decision rule is perceived to be prone to error, the less likely people are to use it, even if using the rule is still better than chance. This supports Dawes’ (1979) hypothesis that people resist mechanical judgment at least partially as a result of the perceived error, or imperfection, of that form of judgment. Still, even when the decision aid’s accuracy rate was stated to be 90%, only 55.3% of Powell’s participants used it. From a statistical point of view, 90% accuracy is extremely impressive; in the area of selection research, it is almost impossible. Shocking is how few participants in the study actually used such a valid statistical rule. Powell suggested that there may be such a thing as an error threshold, or a certain level of error that people are willing to tolerate. Given his findings, it is possible that people cannot tolerate any obvious error—or that the tolerable error threshold may be closer to 100% accuracy, or no error at all.

To test the idea that people cannot tolerate reported error, Kaplan, Reneau, and Whitecotton (2001) manipulated whether predictive ability, or accuracy, of a decision rule was available during a decision task. Ninety-one auditors were asked to make bond rating predictions for 16 corporations based on financial accounting ratios. The results of the study indicated that auditors were significantly more likely to rely on a decision aid when its predictive validity was
unknown, than when it was known: 61% reliance in the known condition versus 79% reliance in
the unknown condition. Kaplan et al. (2001) concluded that the salience of the decision rule’s
error in the known validity condition discouraged participants from using it. When asked to
estimate the accuracy of the decision aid’s predictions, people in the unknown condition
estimated it to be significantly higher than those in the known condition. Although the difference
was not significant, among people in the unknown condition the estimated accuracy of the
decision aid was rated higher than estimated accuracy of one’s own predictions. The authors
speculated that when deciding whether to use a statistical aid, people compare the predictive
ability of the aid to their perceptions of their own abilities. Ostensibly, people believe decision
aids are more prone to error than they themselves are—an idea compatible with Meehl’s (1986)
computer phobia, the overconfidence hypothesis and Dawes’ (1979) idea of cognitive conceit.
It may be that people’s attempt to avoid all error is in part responsible for the reliance on a
particular form of judgment.

One of the reasons people may prefer using holistic judgment is because it is perceived to
be less prone to error than mechanical judgment. Thus, given people’s resistance to error in
prediction (Kaplan et al., 2000; Powell, 1991), I would expect a main effect for amount of error
on attitudes toward judgment approaches. Also, given Powell (1991) and Kaplan et al.’s (2000)
findings that error relates to the resistance to mechanical judgment and that there is a general
preference for holistic judgment, I expect there will be an interaction effect of error rate on
judgment approach, such that the magnitude of the difference in participant reactions to the
selection process will be greater among participants exposed to the mechanical judgment
approach than among participants exposed to the holistic judgment approach. In other words,
error rate will have a stronger negative impact on attitudes toward a selection procedure using mechanical judgment than it will have on attitudes toward a procedure using holistic judgment.

Hypothesis 1a: When error rate is high people will report lower perceived usefulness of the selection procedure than when error rate is low.

Hypothesis 1b: There will be an interaction effect of error rate and judgment approach such that scores on a measure of perceived usefulness will be more negatively affected by differences in error rate for mechanical judgment than they will be for holistic judgment.

Understanding Reactions to Error

In their comprehensive monograph on statistical literacy in medical settings, Gigerenzer and colleagues (2007) asserted that people’s inability to live with uncertainty contributes to the widespread statistical illiteracy and poor decision-making. The authors emphasized that once people “understand that there is no certainty and no zero-risk, but only risks that are more or less acceptable” will people appropriately interpret statistics (p. 58).

In situations where prediction is required, Einhorn (1986) similarly suggested that if people more readily accepted the uncertainty inherent in judgment tasks and the inevitability of some level of error, they would be more likely to use decision aids. Error Management Theory (EMT) is useful for understanding not only why people resist mechanical judgment, but also how knowledge of error and error costs may affect attitudes toward judgment approaches. This is discussed in the following section.

Defining error. In any decision that involves a level of error, there are two types that need to be differentiated: false positive and false negative error. False positive error is when something is predicted to be true, but turns out to be false (e.g. convicting the innocent). False
negative error is when something is predicted to be false but it turns out to be true (e.g. acquitting the guilty).

Although it is not possible to avoid all error in prediction, it may be that people react differently to certain types of error. In the area of legal research, Arkes and Mellers (2002) asked participants to indicate the largest percentage of incorrect convictions (or incorrect acquittals) that society should accept in order to function, given that it is impossible to be correct 100% of the time. They found that most participants were less willing to tolerate false positives (wrong conviction) than false negatives (acquitting the guilty). Whereas the median response for acceptable false convictions was 5%, the median response for an acceptable false acquittal rate was a bit higher, at 8%. Across a variety of situations, EMT provides one perspective for why we observe disproportionate acceptance of false positive versus false negative error.

**Error Management Theory.** Error is inevitable when people make judgments in complex situations filled with “noise” and “uncertainty,” making all forms of judgment under uncertainty prone to error (Haselton & Nettle, 2006). To understand how human systems should react to the inevitability of error, Haselton and Nettle (2006) offered Error Management Theory. Rooted in, and applying the principles of signal detection theory (Green & Swets, 1966; Swets, Dawes, & Monohan, 2001), EMT explains how natural selection influences psychological adaptations to judgment under uncertainty. Specifically, EMT predicts that if judgments are made under uncertainty, and the cost of false negative and false positive errors have been asymmetrical over evolutionary time, natural selection should have a bias toward making judgments in such a way that would result in the least costly error. In other words, in certain situations, biased reasoning strategies may do better than unbiased strategies.
Haselton and Nettle (2006) asserted that when the costs of errors are asymmetrical (e.g., the possibility of death as a result of a false negative mammography result versus the possibility of increased anxiety as a result of a false positive result), “humanly engineered systems” should be biased in the direction of making the less “costly” errors (e.g., anxiety). Although this bias may sometimes increase the overall number of errors, it minimizes the more costly error, and thus minimizes overall cost (Green & Swets, 1966; Swets et al., 2000). The overall increase in error is acceptable here, because certain decision-making adaptations evolved through natural selection to make predictable errors. These adaptations evolved over time as a result of an existing asymmetry between the two types of error, where the accepted error became part of a mechanism that was fashioned toward committing error that is less costly in “reproductive currency” (Haselton & Nettle, 2006). Accordingly, depending on the situation, EMT predicts that human psychology has also evolved with decision rules that are biased toward accepting one type of error over the other.

**Varying costs of error.** In their meta-analytic study on clinical (holistic) versus mechanical prediction, Grove et al. (2001) confirmed that among the varied costs associated with prediction (cost of acquiring information and the cost of making the decision) is the cost of various errors that may result. The cost of error may depend both on the type of error that has occurred, whether positive or negative, and the situation of interest.

From an error management perspective, when it comes to theory development and hypothesis testing, for instance, Type I error is considered as more costly. In such situations, decision-making systems are biased toward making Type II error (Haselton & Nettle, 2006; Luce & Kahn, 1999), thereby reducing the probability of making a Type I error. This allows for more
accurate conclusions when testing hypotheses and developing theories. The reverse approach to
error making is true in high-stake situations, like hazard detection.

In hazard detection, Type II errors, or false negatives are far more dangerous. Imagine a
smoke alarm that does not go off during a fire or a cancer screening which misses a malignant
tumor. With possible death as an outcome, Type II error becomes much more costly. Because of
the serious consequences of not detecting maladies or other types of dangerous situations,
compared to further testing or over-prescription of a treatment, these tests favor false positives.
Overall, people accept this type of error because, in such situations, it is less costly than false
negative error.

Gigerenzer et al. (2007) suggested that for people to make an informed decision based on
health statistics they need to be aware of the fact that screening tests may have benefits and
harms. Benefits include the possibility of identifying disease earlier, where less invasive and
possibly more effective treatments can be used. Harms include monetary costs, inconvenience,
over-diagnosis, and false alarms (i.e., false positive errors). When people make decisions based
on predictions, understanding the cost of a prediction error and its impact on people is critical.
Luce and Kahn (1999) explained that in the health arena, although the cost of false negative error
is clear (e.g., death as a result of overlooking), the effects of false positives are not as easily
understood. Understanding the psychological costs of false positive error has been described as
one of the greatest areas of concern, even though limited evidence is available to judge the
magnitude of the effect (Wardle & Pope, 1982).

In a study of over 300,000 women, for example, survey results indicated that women
receiving false positive mammography test results had more elevated distress and anxiety levels
than women not receiving such results (Brewer, Salz, & Lillie, 2007). Similarly, months after
receiving false positive mammography results, Lerman and colleagues (1991) found that 50% of women reported considerable anxiety about mammograms and breast cancer, and 25% reported that this anxiety affected their daily mood and functioning. In a series of four laboratory experiments, Luce and Kahn (1999) found that people receiving false-positive medical test results (as opposed to negative results) had an increased sense of perceived vulnerability to the relevant malady and believed the relevant test to be less accurate, even when overall false positive rates are known.

Baron and Hershey (1988) suggested that people think about false positives and false negatives separately when the cost of each type of error is asymmetrical. The authors researched the relative weight participants place on prior probabilities, error costs, and test accuracy when deciding whether to conduct or forego a medical test. A series of three experiments showed that participants over-attended to error costs relative to prior probabilities and test accuracy. When evaluating a test, participants also did not seem to comprehend that high prior probabilities make hit rates more relevant, while low prior probabilities make false-positives (false-alarms) more relevant. Previous findings from other tasks have found similar results; people tend to neglect prior probabilities when asked to predict posterior probabilities, referred to as base-rate neglect (Bar-Hillel, 1980; Kahneman & Tversky, 1972). Hamm and Smith (1998), for example, asked patients in a clinic about standard tests for diseases such as strep throat, HIV, and acute myocardial infarction. Each patient was asked to judge 1) the base rate, or the probability that a person has the disease before being tested, 2) the probability that a person tests positive if the disease is present (sensitivity), 3) the probability that a person tests negative if the disease is not present (specificity), and 4) the probability that a person has the disease if test results are positive. Unfortunately, most patients ignored base rate and test accuracy information and judged
all four probabilities to be about the same. Other studies using student participants have also shown that people have a hard time drawing conclusions from sensitivities and specificities (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). In Baron and Hershey’s (1988) study, participants did, however, attend to the fact that a large cost of not treating diseased patients makes a hit rate more relevant, while a large cost for treating non-diseased patients makes false positive rates more relevant. Thus, although people may draw inaccurate conclusions from statistical information and ignore valuable information such as base-rate and hit-rate, they do attend to error costs.

Baron and Hershey (1988) were also able to isolate the role of error costs in people’s choices between two alternative actions. In two of their three experiments the authors asked participants to either choose between two tests (Experiment 1) or choose between administering a test and doing nothing (Experiment 3). In making a decision between two courses of action, the authors found that participants over-attended to cost. Specifically, when asked to choose between two tests, participants chose the test that, should error occur, would result in the less costly error. Additionally, when asked to choose between administering a test and withholding it, participants again chose the approach that would result in the less costly error. Generally, participants based their decision between two approaches on what would be less harmful if error occurred, ignoring all other information (i.e., the respective probability of each error actually occurring). The findings suggest that the cost of error is a critical attribute in this decision process. The authors summarized this finding by stating, “Subjects may sometimes translate costs into a kind of lexical ordering in which minimizing the probability of the worst harm takes priority over all else” (Baron & Hershey, 1988, p. 278). In other words, participants seemed to focus on the
course of action that would minimize the cost of error to patients, even when other statistical information was more relevant.

**Cost of error in selection.** In an attempt to better understand why people prefer holistic judgment in selection, Phillips and Gully (2008) asserted that managers want to make good hiring decisions and if they were more cognizant of the inadequacy of holistic judgment in predicting performance they would not be as likely to rely on it. Similarly, Martin (2008) stated that managers should want to err on the side of fewer false positives, limiting unqualified hires. These authors suggested that hiring the wrong candidate is significantly more noteworthy than overlooking good candidates, implying that managers may want to reduce the probability of false positive error.

Fisher (2008) argued that selection failures, or false positive decisions, should be powerful learning experiences for managers. Hiring unqualified candidates is not a good practice in the vast majority of situations; but whether people believe this to be the case is one of the questions of interest in this research. Because the cost of false positive versus false negative error is asymmetrical in the selection context, we can predict which error would be perceived as more or less acceptable. Furthermore, by describing the error costs we can better isolate how false positive versus false negative costs differentially affect attitudes toward holistic and mechanical judgment approaches.

To better understand the costs of false positive error in selection contexts, it may be helpful to examine high-stake selection situations. Police officers, security personnel, surgeons, lawyers, etc. have an indisputable impact on the lives of other people. Schmidt and Hunter (2000) related several real-word examples in which the failure to use a specific selection decision aid—a cognitive ability test—in selecting employees resulted in a significant decline in
performance. One such example noted by Schmidt and Hunter was of the police force in Washington, D.C., which eliminated the use of cognitive ability tests and background checks. As a result of this change, 80% of the candidates were not able to complete the required training. Consequently, after the training requirements were lowered, there was an increase in firearm accidents, complaints of police abuse and incompetence, and crime on the police force. As Martin (2008) suggested, and as this example illustrates, in selection, avoiding false positive error seems more important than avoiding false negative error. Thus, I suggest there will be a main effect of error type such that attitudes toward the selection procedure will be more favorable when false negative error is emphasized than when false positive error is emphasized.

Hypothesis 2a: When the risk of false negative error is primed people will report higher perceived usefulness of the selection procedure than when the risk of false positive error is primed.

Although error costs seem to be an important factor when making decisions based on predictions reached using certain judgment approaches, there has not been a test of the influences of false positive and false negative error on attitudes toward judgment approaches. If asked to examine the merits of the two judgment approaches, how would the type of error affect people’s perceptions? I would expect false positive and false negative error to differentially affect attitudes toward holistic versus mechanical judgment.

I suggest that the general preference for type of selection procedure (holistic versus mechanical) will interact with reactions to error type. Specifically, because false positive error should be perceived less favorably than false negative error and because attitudes toward mechanical judgment may be more susceptible to the severity of false positive error, given people’s general resistance to it, I propose that false positive error will have a greater negative
impact on attitudes toward selection processes when mechanical judgment is used than when holistic judgment is used.

Hypothesis 2b: There will be an interaction effect of error type and judgment approach such that scores on a measure of perceived usefulness will be more negatively affected by false positive error for mechanical judgment than for holistic judgment.

**Present Study**

The overwhelming evidence in the literature favors mechanical judgment over holistic judgment when making selection predictions. The majority of studies continue to compare holistic to mechanical judgment, while no research has examined how the type of selection error combined with the risk of error (error rate) may impact attitudes toward these two judgment approaches. In this dissertation I empirically studied the effects of judgment approach, error rate and type of error on perceptions of usefulness of a hiring process. This study sampled individuals using a fully crossed 2 (judgment approach: holistic versus mechanical) X 2 (error rate: 10% versus 40%) X 2 (type of error: false positive versus false negative) between-subjects design. After being randomly assigned to one of eight conditions, participants were asked to read a scenario describing a new selection procedure and the possible error that may occur during such a procedure. The participants then responded to the dependant variable measure, perceived usefulness. Diab et al. (2009) also found that participants perceived mechanical selection procedures to be less legal than holistic procedures; thus, perceived legality of the selection procedure was examined as an exploratory dependant variable.
CHAPTER II: METHOD

Participants

Participants were recruited using the StudyResponse Project, a non-profit, academic service that facilitates online research for behavioral, social, and organizational science researchers (Stanton & Weiss, 2002). StudyResponse recruits adult participants via e-mail participation requests. As of 2005, StudyResponse Project had approximately 95,000 people registered in their database. Participation in the study was voluntary but all participants were paid $5.00. StudyResponse contacted a total of 520 individuals; 398 people responded, for a response rate of 76.5%. Further data cleaning removed 4 individuals for missing or miscoded data, 13 for lack of response variations (i.e., answering all 5’s), and 58 individuals for incorrectly answering a check on whether people distinguished between false-positive and false-negative error.\(^1\) The final sample was 323. For a 2 x 2 x 2 design a sample of 250 would achieve power of .80 with an expectation of a medium effect size (Faul, Erdfelder, Lang, & Buchner, 2007); given the sample size in this study, power of .95 was achieved.

Demographic information is summarized in Table 1. Approximately 57% of the sample was female, 83% was white with the remaining participants identifying as African-American (5%), Hispanic (5%), Asian (5%), and “other” (1.5%). All participants were U.S. residents and worked full-time. Participants ranged in age from 21 to 73 with an average age of 40.42 (SD = 10.38). This sample represented 35 different occupation types with the most common being Administrative Support (11%), Education/Training (8%), Managerial (8%), Health or Safety (7%), and Technology (7%).

\(^1\) The item was, “In the scenario you read, what type of hiring mistake could occur?” The response options were, “Someone was hired who turns out to be a bad manager,” and “Someone you did NOT hire would have been a good manager.”
Stimuli and Procedure

Participants were directed to a website where they were randomly assigned to one of eight conditions. They were presented with survey materials and asked to imagine that they are in the role of a head of Human Resources for a large organization looking to implement a new hiring procedure (Appendix A). The scenario began with an introductory paragraph:

Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

The following section of the scenario presented the judgment approach (holistic versus mechanical) independent variable. The holistic procedure read:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, the senior manager will use his own experience and judgment to assess which applicants will be successful managers.

The mechanical procedure read:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, his interview scores will be plugged into a formula that will mechanically combine the scores for an overall assessment of which applicants will be successful managers.
The second independent variable, error rate, has two levels, 10% error versus 40% error. These were chosen to represent low and high levels of error, respectively. The 40% error rate was chosen because it is distinctly higher than what would be interpreted as low or medium error, but not so high as to be perceived on par with chance, or 50% error. The third independent variable, type of error (false positive versus false negative), was presented in conjunction with the level of error. For the false positive condition, with 10% error rate, the scenario continued as follows:

It is important to recognize that this new hiring procedure will not be free of hiring mistakes. It is likely that 1 out of 10 people you hire will not be successful after you hire them. In other words, 10% of the people you hire will not turn out to be good managers for the company.

For the false negative condition, at 40% error, the scenario read:

It is important to recognize that this new hiring procedure will not be free of hiring mistakes. It is likely that 4 out of 10 people you reject would have been successful if you had hired them. In other words, 40% of the people you deny jobs would have turned out to be good managers for the company.

The participants then responded to a series of items designed to elicit reactions to the selection procedure.

**Measures**

**Perceived Usefulness.** Usefulness of the hiring procedure was measured with 8 items. Two items were adapted from the 7-item Perceived Usefulness scale used in Diab et al. (2009) and 6 items were developed for this research. The questions assessed people’s perceptions about the hiring procedure and how useful the procedure is for hiring employees. A sample item is: “I think this procedure will help me make hiring decisions.” Participants responded using a 5-point
Likert scale (1 = strongly disagree to 5 = strongly agree). The internal consistency for the 8-item Perceived Usefulness scale was $\alpha = .92$. Please refer to Appendix B for items.

**Perceived Legality.** Perceived legality of the selection procedure was measured as an exploratory dependent variable. It consisted of three items, two of which were developed by Diab et al. (2009). A sample item is: “I have doubts about the legality of this hiring procedure.” Participants responded using a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). The internal consistency for the Perceived Legality scale was $\alpha = .89$. Please refer to Appendix C for items.
CHAPTER III: RESULTS

Preliminary Analyses

Because the majority of items from Perceived Usefulness and one item from Perceived Legality were developed for this study, exploratory factor analysis was first conducted to confirm the presence of two distinct factors. A principle axis factor analysis, with a varimax rotation, was performed, and two distinct factors were identified. All Perceived Usefulness items loaded on Factor 1 and explained 53.11% of the variance (eigenvalue = 5.84). All Perceived Legality items loaded on Factor 2 and explained 17.50% of the variance (eigenvalue = 1.93). All items had a loading well above .32, the minimum recommended by Tabachnik and Fidel (2001) for retention in a scale (See Table 2).

Levene’s Test of Equality of Variance was performed as a first step to ascertain that homogeneity of variance was not violated. Results supported the assumption of homogeneity of variance for perceived usefulness, $F(7, 315) = 1.13, p = .34$, and for perceived legality, $F(7, 315) = .53, p = .81$.

Hypothesis Testing

Table 3 presents overall means, standards deviations, reliabilities, and intercorrelations for all study variables. A significant correlation was identified between age and Perceived Usefulness, $r = -.17, p < .01$. Consequently, age was controlled for in the analysis as a covariate. To test all hypotheses a 2 x 2 x 2 univariate analysis of covariance was run with Perceived Usefulness as the dependent variable. After controlling for age, a significant main effect was found for judgment approach, $F(1, 314) = 16.13, p < .01, \eta^2 = .05$ (Table 4). Specifically, participants in the holistic condition perceived the selection procedure to be significantly more

---

2 Results for hypothesis testing did not change when a regular univariate analysis of variance (ANOVA) was conducted (i.e., not controlling for age).
useful ($M = 3.13, SD = .76$) than participants in the mechanical condition ($M = 2.77, SD = .82$).
The basis of this research is grounded on the assumption that people generally prefer holistic
judgment methods. The hypotheses proposed, therefore, elaborate on this result, which supports
previous research findings.

Hypothesis 1a stated that people would report lower perceived usefulness of the selection
procedure when the risk of error was high. There was a main effect of error rate, $F(1, 312) =
12.43, p < .01, \eta^2 = .04$, such that the new selection approach was rated significantly more useful
when the error rate was 10% ($M = 3.11, SD = .81$) than when the error rate was 40% ($M = 2.78,$
$SD = .77$). There was no support for Hypothesis 1b, which posited an interaction effect of
judgment approach and error rate (Table 4). In other words, there was no difference in
perceptions of usefulness when the judgment approach was mechanical (versus holistic) and
when the error rate was high (Figure 1). Table 5 presents marginal and cell means and standard
deviations for the perceived usefulness dependent variable.

Hypothesis 2a stated that people would report lower perceived usefulness when the risk
of false positive error, as opposed to false negative error, was emphasized. There was no
significant difference in perceptions of usefulness when the risk of false positive error was
primed ($M = 2.97, SD = .87$) compared to when false negative error was primed, $M = 2.93,$
$SD = .74, F(1, 312) = .62, p > .05, \eta^2 = .00$, not supporting Hypothesis 2a. Likewise, as shown in
Table 4, the predicted interaction of judgment approach and error type was not supported. That
is, when false positive error was emphasized, people’s perceptions of usefulness of mechanical
selection procedures were not more negative than their perceptions of holistic procedures (Figure
2).
Exploratory Analyses

Perceived Legality was used as an exploratory dependent variable. A significant correlation was found between Perceived Legality and education level of participants, $r = .16$, $p < .01$. A univariate analysis of covariance, with education level as the covariate and Perceived Legality as the dependent variable was performed\(^3\). As shown in Table 6, significant main effects of judgment approach and error type were identified. Specifically, participants in the holistic condition perceived the selection procedure to be significantly more legal ($M = 3.28$, $SD = .91$) than participants in the mechanical condition, $M = 3.08$, $SD = .97$, $F(1, 314) = 6.04$, $p < .05$, $\eta^2 = .02$. Additionally, the ratings of perceived legality were significantly lower in the false negative condition ($M = 3.05$, $SD = .95$), than in the false positive condition, $M = 3.30$, $SD = .92$, $F(1, 314) = 5.95$, $p < .05$, $\eta^2 = .02$. Concretely, the selection procedure described as leading to false negative error was perceived to be significantly less legal than the procedure described as leading to false positive error. Table 7 presents marginal and cell means and standard deviations for perceived legality.

\(^3\) Results for exploratory analyses did not change when a regular univariate analysis of variance (ANOVA) was conducted (i.e., not controlling for education).
CHAPTER IV: DISCUSSION

The primary contribution of this study is the consideration of the role of error in people’s attitudes toward holistic versus mechanical judgment in employee selection procedures. Error rate and type of error (false positive versus false negative) were manipulated to determine the influence these two factors have on perceptions of usefulness and perceptions of legality (exploratory variable). Results from this study support extant research findings that people prefer holistic over mechanical judgment approaches (Corey & Merenstein, 1987; Hilton & Simmons, 2001; Lievens, et al., 2005; Terpstra, 1996). Focusing solely on the hypothesized outcomes, it was surprising to find that manipulation of error type did not shed light on this preference for holistic over mechanical judgment. Results of the experimental analyses, however, do provide some insight into how type of error may influence people’s attitudes toward selection procedures.

First considering error rate, as hypothesized, people reported lower perceived usefulness of the selection procedure when error rate was high (40% error) than when error rate was low (10% error). The sensitivity to error rates found in this study is noteworthy because previous findings indicated that people only marginally attend to error rate information. Powell (1991), for instance, found that although utilization of a mechanical rule increased at higher levels of reported accuracy, even at the highest level of accuracy (90%, versus 50% and 70%) only 55% of participants utilized the rule. Additionally, Hamm and Smith (1998) found that people tended to ignore important statistical information like test accuracy when making health judgments. In effect, people judged all four probabilities asked about in their study to be about the same. Baron and Hershey (1988) similarly found that people overweighed subjective cost of the two types of errors relative to test accuracy and prior probabilities. As the significant manipulation of error
rate indicates, participants in this study did consider the probability information in making their assessments of perceived usefulness.

It was also proposed that attitudes toward a mechanical selection procedure described as having a high error rate would be more negatively affected than attitudes toward a holistic procedure described with the same level of error. This prediction stemmed from research showing that people generally resist mechanical judgment in favor of more holistic approaches. The combination of a seemingly universal resistance to mechanical judgment and a desire to achieve near perfect prediction would suggest that this combination of an undesirable procedure with a high level of error would appear disproportionately worse than the combination of a desirable procedure with a similar level of error. The results of this study, however, did not support this proposition, as there was not a significant interaction of judgment approach and error rate; the differences in perceptions of usefulness across both procedures were approximately equivalent when comparing procedures described with 10% error and procedures described with 40% error.

Unlike the non-significant interaction effect, the main effect of error rate does contribute to our expanding understanding of people’s preferences toward selection procedures. These findings tell us that not only are people sensitive to differences in error rate, but that they also (appropriately) judge a selection procedure more prone to error to be less valuable than one that is less prone to error. In other words, people seem to consider error in their judgments of usefulness. What is particularly notable here is that unlike in previous research, such as when people ignored error information in a basic predictive judgment task (Powell, 1991), as well as under-weighed error in various health-related judgment scenarios (Baron & Hershey, 1988; Hamm & Smith, 1998), that was not the case in this employee selection study. As we learn more
about people’s perceptions of different judgment approaches, and how these approaches define and differentiate selection procedures, it may be helpful to know that less error may simply be considered better than more error. As Dawes (1979) suggested three decades ago, people resist using those forms of judgment that are perceived to be prone to error. Overall, a desire to avoid, or minimize error may have a direct relation to people’s perceptions of the value of a selection procedure. In the future, it may also be helpful to explore boundary conditions around which situations do and do not elicit this error rate effect. Namely, the question may be whether there is something about the employee selection situation that supports an error rate effect more so than other types of scenarios (e.g., medical).

In addition to error rate, this study also examined the effect of error type on people’s perception of holistic versus mechanical judgment. Error Management Theory, proposed by Haselton and Nettle (2006), offers that when the cost of error is asymmetrical, decision-making is biased in the direction of making the least costly error. The authors provided numerous examples demonstrating how EMT may explain biases in survival-based perception, attention, and learning, as well as biases in social and self-perception. Based on this theory, the second hypothesis posited that people would perceive the cost of false positive error to be more serious than the cost of false negative error. Additionally, an interaction effect was predicted, such that when false positive error risks were described, people’s perceptions of usefulness of mechanical selection procedures would be more negative than perceptions of holistic procedures. Even though practitioners and academics agree that the consequences of false positive error in selection are more costly (Fisher, 2008; Martin, 2008; Phillips & Gully, 2008; Schmidt & Hunter, 1998), there was no significant effect of error type. The difference in perceived usefulness ratings between the selection procedure described as leading to false positive error
and the procedure described as leading to false negative error was trivial; neither the hypothesized main effect nor the interaction were supported.

Although it may seem clear that one type of error would lead to more significant costs than the other, one possible reason for the lack of an error type effect is that lay people may not perceive that asymmetry. Luce and Kahn’s (1999) research, for instance, suggested that the two error types were almost identically associated with the general concept of test inaccuracy. These findings were surprising given that the authors expected participants to differentiate between false positive error and false negative error. The lack of a perceived distinction carries important implications for how people think about types of error. In the present study there was an expectation of asymmetry in perceived cost of error, which was not supported by the results. Given Luce and Khan’s findings, however, it is possible that in the context of employee selection people may not consider one type of error to be significantly more costly than the other.

It should be noted that failing to find a significant result can not rule out the possibility that costs of false positive and false negative error truly are perceived to be asymmetrical. It may be that the dependent variable in this study, perceived usefulness, is not the ideal measure for these types of costs. A more direct test of perceptions of false positive versus false negative error costs should be considered before making the determination that “error is just error.” Directly asking participants about which error is perceived to be better or worse may be a more appropriate start to making the asymmetry argument. In a within-subjects design, for instance, participants could be asked to rate and/or rank false positive and false negative errors and their related costs. Participants could additionally be asked to explain their responses, providing further insight into their beliefs about the different types of error; one error may be more socially or morally acceptable than the other, for example, an idea I also return to later.
Exploratory Analyses: Perceived Legality

Motivating the inclusion of a perceived legality scale in this study was the finding by Diab et al. (2009) that mechanical combination of interview scores was perceived to be less legal than holistic combination of interview scores. Replicating Diab et al.’s results, the participants in this research also reported the mechanical selection procedure to be less legal than the holistic procedure, again adding to the growing body of literature supporting the idea that people generally prefer intuitive judgment approaches. In their own exploratory analyses, Diab and colleagues additionally found that people perceived mechanical judgment to be “insufficient” compared to holisitic judgment, a belief which has been echoed by highly educated academicians (Arkes, 2008). These results may suggest that beliefs about the (in)sufficiency of a judgment approach influence perceptions of its legality.

If insufficient procedures are perceived to be less legal, then people may resist using mechanical judgment because they fear some form of reprisal, legal or otherwise. In the medical context, for instance, Arkes and colleagues (2007) offered that one reason why physicians resist the use of mechanical decision aids is fear that such aids might increase the probability of having an adverse verdict in a malpractice trial (cf., Arkes, Shaffer, & Medow, 2008). Given that these authors also found that patients judged physicians, as well as the overall satisfaction with their medical visit more negatively when the physicians used a mechanical decision aid in making diagnoses, as opposed to no aid (i.e., relying on professional expertise), lends substance to this supposition.

In the case of the error rate manipulation, results with perceived legality were more ambiguous than with the judgment approach manipulation. Although observed differences were

---

4 Insufficient in Diab et al. (2009) was defined by the following items: inadequate, insufficient, imprecise, deficient, and incomplete.
in the expected direction, there was no significant effect of error rate on perceived legality. Given the significant effect of error rate on perceived usefulness, this finding may suggest that the extent of error in a procedure does not influence how legal it is perceived to be. There was a significant effect of error type on perceived legality, however, indicating that people may be more concerned with the type of procedure and type of error when judging legality of a selection approach, not necessarily the extent of error.

Examining the effect of error type then, false negative error was perceived to be less legal than false positive error. From this result, we can infer that people may be more concerned with the repercussions of rejecting candidates than they are with the repercussions of employing unsuccessful candidates. If one is to presume that judgments of legality are an extension of judgments of ethical and/or moral behavior (i.e., what is right versus what is wrong), then the significant effect of error type could be a reflection of societal values. Arkes and Mellers (2002), for example, found that people are more tolerant of acquitting the guilty than they are of convicting the innocent. In a similar vein, these results indicate that people are more tolerant of employing the unsuccessful than they are of rejecting the successful. Research on omission bias has repeatedly demonstrated people’s preference for harms of omission over harms of action (Baron & Ritov, 2004; Ritov & Baron, 1990). Spranca, Minsk, and Baron (1991) found that even with intentions held constant, many participants considered commissions that caused harm to be morally worse than omissions that caused harm, a bias that may explain the current results. Actively deciding to implement a selection procedure that explicitly indicates successful applicants will be rejected may be perceived as directly causing harm, a situation that many people would prefer to avoid.
Baron, Bazerman, and Shonk (1996) presented a selection of biases that they asserted best account for current sub-optimal practices in social decision making, and that point to the idea that people resist the prudent trade-off of accepting small losses in exchange for larger gains. Given the extensive research indicating that people are often not able to make wise statistically based judgments (Baron & Hershey, 1988; Cosmides & Tooby; Gigerenzer & Hoffrage, 1995; Gigerenzer et al., 2005), this idea is not new to the field of judgment and decision making. Over four centuries ago Arnauld argued, “We must enlighten those persons who take extreme and vexatious precautions for the preservation of life and health by showing that these precautions are a much greater evil than is the remote danger of the mishaps feared” (Arnauld, 1662/1964, p. 356). Einhorn (1986) echoed Arnauld and suggested that in situations of prediction, people need to accept some level of error in order to reduce overall error. In terms of an employee selection procedure that has a less than perfect correlation with the criterion, people may need to be more accepting of false negatives in order to optimize the overall usefulness of the procedure. Further investigation into why people perceive false negative error to be less legal than false positive error may shed light on how to best communicate the value of less than perfect prediction. It is quite probable that people may believe a selection procedure that rejects qualified candidates is inappropriate, not knowing that with such a procedure the likelihood of selecting不合格 candidates is also reduced. How people think about these two types of error may help us understand how best to bridge the gap in improving acceptance of false negative error.

Limitations

When considering the generalizability of these findings, several potential limitations should be noted. First, although this study instructed participants to place themselves in the role
of an HR leader, there is no assurance that participants were not also thinking about the selection procedure from a more personal point of view, specifically, that of a possible applicant. Consequently, the results here may be more indicative of applicant reactions, than of HR professionals’ beliefs. The question that follows then, is does the point of view matter? Diab, et al. (2009), in an initial attempt to examine lay perceptions of decision aids found that perceptions were consistent across people, whether they were placed in the role of employer or applicant. Given Diab et al.’s findings, it may be that participants in the current study considered the implications of the selection procedure from various points of view, not just as an HR leader, possibly neutralizing the manipulation.

Accordingly, a second limitation, and one particularly important for the null result of error type, is the possibility that the manipulation was not strong enough. For there to be bias toward one type of error over the other, there must be a perceived asymmetry of error costs (Haselton & Nettle, 2006). Because the descriptions of false positive and false negative error were very straightforward one-sentence definitions, they may not have elicited the necessary perception in asymmetry of cost. Unambiguous examples of each type of error may more effectively prime participants to focus on the asymmetry of costs, which might not have been possible with the current textbook definitions. Future studies, therefore, should explore using scenarios that plainly state the cost of false positive error and false negative error (e.g., dollars lost by an organization as a result of false positive error or the missed opportunity to hire an extraordinary candidate as a result of false negative error) and not put the participant in the position of needing to make that implied connection on their own.

In the medical arena, research has suggested that one reason why people may not comply with appropriate preventative treatments or screening tests is because the consequences of
noncompliance are difficult to imagine (Sherman, Cialdini, Schwartzman, & Reynold, 1985). People tend to feel more vulnerable, and are thus more likely to comply, when consequences of a disease are described more vividly (Leventhal 1970; Rogers & Mewborn, 1976) and with cognitive elaboration (McGill & Anand, 1989). In the case of the present null results, a stronger and more vivid emphasis on the actual consequences of each type of error may be necessary for people to perceive the asymmetry in error costs. When deciding on a course of action, Baron and Hershey (1988) repeatedly found that people ignored vital statistical information in favor of avoiding the most costly error, implying that people are sensitive to differences in error outcomes. And although an item was used to ensure participants attended to the error type manipulation in this study, it only verified whether participants noted the type of hiring mistake that was primed and not the implications of that hiring mistake. The lack of a transparent and, most importantly salient cost to error may not have allowed for such differences to be identified in this research.

**Future Research**

Based on the results of this study, there are a number of concerns that can be addressed in future research. First, the lack of interaction effects may suggest that error does not differentially affect judgment approaches. However, the design of the present study does not allow for the question to be answered completely. The idea that people may differentially weigh discrete aspects of a selection procedure is still warranted. Because this study employed a fully crossed between-subjects design, it is not possible to determine which variable—judgment approach, error rate, or error type—is of greater importance to participants. Although both judgment approach and error rate had a main effect on perceptions of usefulness, it is not clear how the *combination* of the two affects perceptions. A policy capturing study, for instance, would be
better suited to answering this question. This was how Lievens et al. (2005) found that across multiple combinations of predictors people placed more weight on holistic, compared to mechanical, selection procedures.

A within-subjects design might also shed more light on how the combination of factors influences attitudes. Exploring the effects of joint versus separate evaluation mode, Brooks et al. (2009) found that holistic hiring practices were perceived more favorably. In joint evaluation mode, in fact, differences between the two approaches were three times as large as they were in separate evaluation mode. The main effects of judgment approach and error rate found in this study indicated that people were concerned with both variables; the lack of an interaction effect still leaves the question of how important each variable is in comparison to the other. Side-by-side comparisons with different levels of error and ratings of different combinations of the two variables may provide more clarity to the present findings.

Recently, the management field has been moving toward “evidence-based management,” which Rousseau (2006) defined as “translating principles based on best evidence into organizational practices” (p. 256). As Rynes and colleagues (2002) demonstrated, however, the same HR executives who are to be moving their organizations into a new era of evidence-based practices are often misinformed about the social science and organizational research findings that Rousseau places in the forefront of evidence-based management. In addition to a clearer understanding of employment law and more transparent risk communication, organizational decision-makers may need the type of education in statistics suggested for medical professionals and patients (Gigerenzer et al., 2007). In light of the current results, future research needs to clarify why false negative error is perceived to be less legal than false positive error. If it is a matter of limited understanding of statistics, as opposed to it being an issue of ethical
appropriateness, then the problem can more easily be addressed. Although not supported by
evidence in his study, Ahlburg (1992) nevertheless returned to the idea that minimizing legal
challenges may be a driving force behind the decision to use or not to use a selection tool.
Therefore, understanding why false negative error is perceived to be less legal than false positive
error seems to be of particular value.

Finally, although this study offers that it is the cost of a particular type of error that drives
whether it is accepted by the individual or not (Haselton & Nettle, 1998), a direct test of this
assertion is necessary. Is one type of error truly perceived to be more costly than the other? And
if there is a perceived asymmetry in error costs, it may be that, contrary to what was proposed in
the introduction, the asymmetry may be in the direction of avoiding overlooking talent as
opposed to minimizing bad hires.

Conclusion

To date, no previous research has examined the role of error rate and type of error in
influencing judgment preferences. Although there was no evidence of a differential effect of
error rate or type of error on attitudes toward the two judgment approaches, both of these factors
did have a direct effect on attitudes. Participants’ perceived the selection procedure described as
having a low error rate as more useful than the selection procedure described with a higher error
rate. More research is needed to determine the relative weight participants place on judgment
approach versus error rate. Participants also perceived the selection procedures emphasizing false
negative error to be less legal than the selection procedure emphasizing false positive error. This
finding begs for future research into the perceived costs of the two types of error. Error
Management Theory offers that when an asymmetry in costs exists, decision strategies will be
biased in the direction of making the least costly error. The question remains, do people perceive
the consequences of false negative errors to be more costly than the consequences of false positive error and vice versa?
REFERENCES

Arkes, H.R. (2008). Being an advocate for linear models of judgment is not an easy life. In  
*Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (Eds. R.M.  

Behavior*, 26, 625-639.

rule in a probabilistic task. *Organizational Behavior and Human Decision Processes*, 37,  
93-110.

vulnerability to malpractice verdicts. Unpublished manuscript.

Arnauld, A. (1964). *The art of thinking (Port Royal Logic)* (J. Dickoff & P. James, Trans.).  
Indianapolis, IN: Bobbs-Merrill. (Original work published 1662)


211-233.

costs, and test accuracy. *Organizational Behavior and Human Decision Processes*, 41,  
259-279.

*Organizational Behavior and Human Decision Processes*, 94, 74–85.


Perspectives on Judgment and Decision Research (pp. 13-63). New York: Cambridge University Press.
APPENDIX A

Stimuli

Holistic – 10% – False Negative

Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, the senior manager will use his own experience and judgment to assess which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 1 in 10 people you reject would have been successful if you had hired them. In other words, 10% of the people you deny jobs would have turned out to be good managers for the company.
Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, the senior manager will use his own experience and judgment to assess which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 4 in 10 people you reject would have been successful if you had hired them. In other words, 40% of the people you deny jobs would have turned out to be good managers for the company.
Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, the senior manager will use his own experience and judgment to assess which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 1 in 10 people you hire will not be successful after you hire them. In other words, 10% of the people you hire will not turn out to be good managers for the company.
Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, the senior manager will use his own experience and judgment to assess which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 4 in 10 people you hire will not be successful after you hire them. In other words, 40% of the people you hire will not turn out to be good managers for the company.
Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, his interview scores will be plugged into a formula that will mechanically combine the scores for an overall assessment of which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 1 in 10 people you reject would have been successful if you had hired them. In other words, 10% of the people you deny jobs would have turned out to be good managers for the company.
Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, his interview scores will be plugged into a formula that will mechanically combine the scores for an overall assessment of which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 4 in 10 people you reject would have been successful if you had hired them. In other words, 40% of the people you deny jobs would have turned out to be good managers for the company.
Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, his interview scores will be plugged into a formula that will mechanically combine the scores for an overall assessment of which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 1 in 10 people you hire will not be successful after you hire them. In other words, 10% of the people you hire will not turn out to be good managers for the company.
Imagine you are in charge of human resources for a large company that is constantly hiring a large number of managers for its various locations around the world. Currently, people are hired based solely on their resumes. But, you have been instructed to implement a new procedure for hiring managers, which will be used along with the resume. We are interested in your reactions to this procedure:

Job applicants will be interviewed by a senior manager and scored on their job knowledge, decision-making ability, and leadership potential.

Following this, his interview scores will be plugged into a formula that will mechanically combine the scores for an overall assessment of which applicants will be successful managers.

It is important to recognize that this new procedure will not be free of hiring mistakes. It is likely that 4 in 10 people you hire will not be successful after you hire them. In other words, 40% of the people you hire will not turn out to be good managers for the company.
APPENDIX B

Perceived Usefulness Scale

1. I think there are better ways to hire people.*
2. I don’t think this is the way to hire an employee.*
3. I would continue to use this hiring procedure.
4. This is a great way to hire people.
5. I think this procedure is useful.
6. I would rather not use this hiring procedure.
7. This is not a good way to hire people.
8. I think this procedure will help me make hiring decisions.

* Items adapted from Diab et al. (2009).
APPENDIX C

Perceived Legality Scale

1. I have my doubts about the legality of this hiring procedure.*
2. I am not sure this is legal.*
3. I could get sued if I use this procedure.

* Items adapted from Diab et al. (2009).
Table 1

Description of Sample

<table>
<thead>
<tr>
<th>Age</th>
<th>40.42 (10.38)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Education Level</td>
</tr>
<tr>
<td>Male</td>
<td>Associates Degree</td>
</tr>
<tr>
<td>Female</td>
<td>Some College</td>
</tr>
<tr>
<td></td>
<td>Bachelors Degree</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>Some Graduate School</td>
</tr>
<tr>
<td>Caucasian</td>
<td>Masters Degree</td>
</tr>
<tr>
<td>African-American</td>
<td>Advanced Degree</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5</td>
</tr>
<tr>
<td>Asian</td>
<td>Administration Support</td>
</tr>
<tr>
<td>Other</td>
<td>Education/Training</td>
</tr>
<tr>
<td></td>
<td>Managerial</td>
</tr>
<tr>
<td></td>
<td>Health or Safety</td>
</tr>
<tr>
<td></td>
<td>Technology</td>
</tr>
</tbody>
</table>

Note. Numbers in columns are percentages with the exception of Age, which is an average, with Standard deviation in parentheses.
Table 2.

*Results of Principal Axis Factoring on Dependent Variables*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think there are better ways to hire people.*</td>
<td>.60</td>
<td>.27</td>
</tr>
<tr>
<td>I don’t think this is the way to hire an employee.*</td>
<td>.71</td>
<td>.26</td>
</tr>
<tr>
<td>I would continue to use this hiring procedure.</td>
<td>.84</td>
<td>.04</td>
</tr>
<tr>
<td>This is a great way to hire people.</td>
<td>.86</td>
<td>.04</td>
</tr>
<tr>
<td>I think this procedure is useful.</td>
<td>.70</td>
<td>.21</td>
</tr>
<tr>
<td>I would rather not use this hiring procedure.*</td>
<td>.78</td>
<td>.21</td>
</tr>
<tr>
<td>This is not a good way to hire people.*</td>
<td>.76</td>
<td>.34</td>
</tr>
<tr>
<td>I think this procedure will help me make hiring decisions.</td>
<td>.73</td>
<td>.17</td>
</tr>
<tr>
<td>I have my doubts about the legality of this hiring procedure.*</td>
<td>.20</td>
<td>.87</td>
</tr>
<tr>
<td>I am not sure this is legal.*</td>
<td>.18</td>
<td>.88</td>
</tr>
<tr>
<td>I could get sued if I use this procedure.*</td>
<td>.18</td>
<td>.78</td>
</tr>
</tbody>
</table>

*Note.* Factor 1 is Perceived Usefulness; Factor 2 is Perceived Legality. * Items were reverse scored for hypothesis testing.
Table 3

*Means, Standard Deviations, and Intercorrelations for Study Variables*

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>40.42</td>
<td>10.38</td>
<td>-.05</td>
<td>-.15</td>
<td>-.01</td>
<td>-.17*</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.57</td>
<td>0.50</td>
<td>--</td>
<td>-.05</td>
<td>-.05</td>
<td>.00</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>1.40</td>
<td>1.03</td>
<td>--</td>
<td>.04</td>
<td>.01</td>
<td>-.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td>5.04</td>
<td>1.40</td>
<td>--</td>
<td>.06</td>
<td>.16*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Usefulness</td>
<td>2.95</td>
<td>0.81</td>
<td>(.92)</td>
<td>.41*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Legality</td>
<td>3.18</td>
<td>0.94</td>
<td>(.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p* < .01; Cronbach’s alpha for Perceived Usefulness and Perceived Legality presented on the diagonal.
Table 4

Summary of Analysis of Covariance: Effects of Judgment Approach, Error Type, and Error Rate on Perceived Usefulness

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1</td>
<td>6.15</td>
<td>10.48*</td>
<td>.03</td>
</tr>
<tr>
<td>Judgment Approach (A)</td>
<td>1</td>
<td>9.46</td>
<td>16.13*</td>
<td>.05</td>
</tr>
<tr>
<td>Error Rate (B)</td>
<td>1</td>
<td>7.29</td>
<td>12.43*</td>
<td>.04</td>
</tr>
<tr>
<td>Error Type (C)</td>
<td>1</td>
<td>.36</td>
<td>.62</td>
<td>.00</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>.19</td>
<td>.33</td>
<td>.00</td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>.12</td>
<td>.20</td>
<td>.00</td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>1.15</td>
<td>1.96</td>
<td>.01</td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>1.17</td>
<td>1.99</td>
<td>.01</td>
</tr>
<tr>
<td>Error</td>
<td>312</td>
<td>.59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .01.
### Table 5

*Marginal and Cell Means and Standard Deviations by Independent Variable for Perceived Usefulness, with Age as a Covariate*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Condition</th>
<th>Judgment</th>
<th>Error Rate</th>
<th>Error Type</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Holistic</strong></td>
<td>1</td>
<td>10%</td>
<td>False Negative</td>
<td>46</td>
<td>3.11</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>False Positive</td>
<td>45</td>
<td>3.40</td>
<td>.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>40%</td>
<td>False Negative</td>
<td>41</td>
<td>3.01</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>False Positive</td>
<td>33</td>
<td>2.93</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>165</td>
<td>3.13</td>
<td>.76</td>
</tr>
<tr>
<td><strong>Mechanical</strong></td>
<td>5</td>
<td>10%</td>
<td>False Negative</td>
<td>32</td>
<td>2.88</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>False Positive</td>
<td>48</td>
<td>2.97</td>
<td>.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>40%</td>
<td>False Negative</td>
<td>31</td>
<td>2.60</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td>False Positive</td>
<td>45</td>
<td>2.58</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>156</td>
<td>2.77</td>
<td>.82</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>171</td>
<td>3.11</td>
<td>.81</td>
</tr>
</tbody>
</table>

*Note:* Marginal means and standard deviations are in boldface.
Table 6

Summary of Analysis of Covariance: Effects of Judgment Approach, Error Type, and Error Rate on Perceived Legality

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education Level</td>
<td>1</td>
<td>7.20</td>
<td>8.50*</td>
<td>.03</td>
</tr>
<tr>
<td>Judgment Approach (A)</td>
<td>1</td>
<td>5.11</td>
<td>6.04*</td>
<td>.02</td>
</tr>
<tr>
<td>Error Rate (B)</td>
<td>1</td>
<td>.56</td>
<td>.66</td>
<td>.00</td>
</tr>
<tr>
<td>Error Type (C)</td>
<td>1</td>
<td>5.04</td>
<td>5.95*</td>
<td>.02</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>.04</td>
<td>.05</td>
<td>.00</td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>1.06</td>
<td>1.26</td>
<td>.00</td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>.98</td>
<td>1.16</td>
<td>.00</td>
</tr>
<tr>
<td>Error</td>
<td>314</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *$p < .01$, **$p < .01$.  

Table 7

Marginal and Cell Means and Standard Deviations by Independent Variable for Perceived Legality, with Education as a Covariate

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Judgment</th>
<th>Condition</th>
<th>Approach</th>
<th>Error Rate</th>
<th>Error Type</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>Holistic</td>
<td>10%</td>
<td>False Negative</td>
<td>46</td>
<td>3.09</td>
<td>.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td>False Positive</td>
<td></td>
<td>46</td>
<td>3.59</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>40%</td>
<td>False Negative</td>
<td>41</td>
<td>3.20</td>
<td>.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td>False Positive</td>
<td></td>
<td>34</td>
<td>3.25</td>
<td>.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>167</td>
<td>3.28</td>
<td>.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>Mechanical</td>
<td>10%</td>
<td>False Negative</td>
<td>32</td>
<td>2.97</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td>False Positive</td>
<td></td>
<td>48</td>
<td>3.22</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td>40%</td>
<td>False Negative</td>
<td>31</td>
<td>2.87</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td>False Positive</td>
<td></td>
<td>45</td>
<td>3.14</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>156</td>
<td>3.07</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>10%</td>
<td></td>
<td>172</td>
<td>3.23</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>40%</strong></td>
<td></td>
<td></td>
<td></td>
<td>151</td>
<td>3.12</td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Negative</td>
<td></td>
<td></td>
<td></td>
<td>150</td>
<td>3.05</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>False Positive</td>
<td></td>
<td></td>
<td></td>
<td>173</td>
<td>3.30</td>
<td>.92</td>
</tr>
</tbody>
</table>

Note: Marginal means and standard deviations are in boldface.
Figure 1. Non-significant interaction of judgment approach and error rate.
Figure 2. Non-significant interaction of judgment approach and error type.