Use Authorization

In presenting this dissertation in partial fulfillment of the requirements for an advanced degree at Idaho State University, I agree that the Library shall make it freely available for inspection. I further state that permission to download and/or print my dissertation for scholarly purposes may be granted by the Dean of the Graduate School, Dean of my academic division, or by the University Librarian. It is understood that any copying or publication of this dissertation for financial gain shall not be allowed without my written permission.

Signature _____

Date _____

THE ACCURACY OF JUDGING RISK IN OTHERS: INVESTIGATING INFORMATION QUALITY AND INDIVIDUAL DIFFERENCES AS MODERATORS OF ACCURACY OF RISK TRAIT JUDGEMENTS

by

Jared Vineyard

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctorate of Philosophy in Experimental Psychology

in the Department of Psychology

Idaho State University

Spring 2016

To the Graduate Faculty:

The members of the committee appointed to examine the dissertation of Jared Vineyard find it satisfactory and recommend that it be accepted.

_____

Tera D. Letzring, Ph.D.

Committee Chair

_____

Steven R. Lawyer, Ph.D.

Committee Member

_____

Xiaomeng Xu, Ph.D.

Committee Member

_____

Nicki Aubuchon-Endsley, Ph.D.

Committee Member

_____

Alex Bolinger, Ph.D.

Graduate Faculty Representative

Idaho State
UNIVERSITY
Office for Research Integrity
921 South 8th Avenue, Stop 8046 • Pocatello, Idaho 83209-8046

July 1, 2015

Vineyard Jared
Psychology
MS 8112

RE: regarding study number IRB-FY2015-66 : Getting to Know Risk: Judging Risk Propensity and Perceptions in Others

Dear Mr. Vineyard:

I have reviewed your request for expedited approval of the new study listed above. This is to confirm that I have approved your application.

Notify the HSC of any adverse events. Serious, unexpected adverse events must be reported in writing within 10 business days.

Submit progress reports on your project in six months. You should report how many subjects have participated in the project and verify that you are following the methods and procedures outlined in your approved protocol. Then, report to the Human Subjects Committee when your project has been completed. Reporting forms are available on-line.

You may conduct your study as described in your application effective immediately. The study is subject to renewal on or before Wed, 29 Jun 2016, unless closed before that date.

Please note that any changes to the study as approved must be promptly reported and approved. Some changes may be approved by expedited review; others require full board review. Contact Tom Bailey (208-282-2179; email humsubj@isu.edu) if you have any questions or require further information.

Sincerely,

Ralph Baergen, PhD, MPH, CIP
Human Subjects Chair

Phone: (208) 282-2179 • Fax: (208) 282-4723 • www.isu.edu/research

DEDICATION

I would like to dedicate this work to my dearest wife Teresa and my two children Genevieve and Carson.

ACKNOWLEDGEMENTS

My gratitude goes out to my committee for their commitment of time and expertise to this project. I would like to recognize especially my advisor Dr. Tera Letzring for taking me under her wing for these last two years of the program. Thank you for your continued support. I would like to thank Dr. Nicki Aubuchon-Endsley for her ongoing support and statistical expertise. My research has been strengthened by your contributions and your steadfast encouragement to persevere. Thank you Dr. Kathleen McCulloch for supporting my application to graduate school in the first place.

The work contained here represents the contributions of and sacrifices of several people across my time in graduate school. I would like to acknowledge my mother Faylene for raising me with the grit necessary to accomplish all I have. To my dearest wife Teresa, all my love, and appreciation for standing by me through these trials. Your unwavering support, love, and compassion are the threads that bound up my resolve and carried us forward. I appreciate all you did during these years including taking care of the children and house while I disappeared into the office for hours at a time. Also, your intellectual contributions to my entire graduate school experience were always welcome, and priceless. Thank you for listening to the endless practice of presentations for class and thesis and letting me bounce my ideas off of your sharp, keen, and wonderfully creative mind. Thank you for all you have done darling, I would not be here without you.

Finally, to my two children Genevieve and Carson. You both are my inspiration, and I could not imagine doing this without you. As a family, we have come this far, and we have nothing but time and our dreams from here on out.

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

Research on personality judgment has explored the accuracy of judgments of broad personality traits in zero aquaintance circumstances, however, little is known about the accuracy of judgments about trait risk components in others. This study investigated the accuracy of judging risk propensity and attitudes in others, and relevant moderators. These moderators included how information relevance and judge's individual risk propensity influence accuracy. Participants completed self-report measures of risk, and then observed one-minute video interviews of seven targets and made judgments of risk propensity and attitudes for each target. The judges were randomly assigned to either watch targets discuss their personality (low risk-relevance), attitudes towards risk (broad risk-relevance), or a particular risky behavior in which they had engaged (specific risk-relevance). It was hypothesized that risk propensity and attitudes would be judged accurately across targets, as information relevance increases accuracy would also increase, and judge's risk level would interact with judge-target ratings and information relevance to produce higher accuracy at high judge-risk levels. Results demonstrated that across risk relevance conditions judges accurately rated targets on risk constructs for both normative and distinctive accuracy. Judges who watched targets discuss their attitudes towards risk had significantly higher distinctive accuracy than the low relevance and specific relevance conditions. Judges who watched targets discuss a risky behavior had lower distinctive accuracy than both the low relevance and broad relevance conditions. Individual differences in judge risk propensity did not moderate accuracy of the relationship between information relevance and accuracy. These finding demonstrate that people can accurately judge trait risk constructs after viewing one-minute video clips of

targets. Findings also support previous research on information relevance and add new insights into how different conversation content can impact accuracy.

Chapter I: Introduction

Every day we make judgments and estimations to fill in incomplete information. For example, when meeting a potential new employee for the first time, the supervisor tries to evaluate his or her personality to estimate and predict future job performance with minimal interaction and information. When we meet a potential new friend for the first time, we try to do the same and judge the characteristics essential to a future friendship. A significant amount of research has investigated and established that people can and do make accurate judgments of others (Ambady, Hallahan, & Rosenthal, 1995; Funder, 1995, 2012). In fact, there is a growing body of literature that demonstrates that first impressions lead to accurate judgments of personality traits in situations where people do not know each other, which is referred to as zero acquaintance (Funder & Colvin, 1988; Hall et al., 2008; Naumann, Vazire, Rentfrow, & Gosling, 2009). While there is evidence to support the accuracy of personality judgments for broad traits (e.g., the Big Five), there is minimal research into the accuracy of judgments for risk propensity and risk attitudes in others.

Risk propensity is the degree to which a person is willing to take chances or embrace a riskier option over a less risky option, while risk attitudes are the evaluations of the likelihood and benefits associated with an outcome (Sitkin & Weingart, 1995; Stewart & Roth, 2001). The ability to estimate risk propensity and attitudes in others has significant social importance such as avoiding threats (Stillman, Maner, & Baumeister, 2010) and identifying social opportunities (Haselton & Funder, 2006). For example, within the context of current threats to privacy and online security, it is important to study and understand how individuals determine who will keep their information safe or act

recklessly with it. As research into online security has suggested, how individuals perceive risk and their overall attitudes towards risk are valuable to understand and warrant further investigation to improve personal and corporate security practices (West, 2008). Also, the vast majority of decisions people make involve some risk and risk propensity is a key trait that underlies critical outcomes across life (Zuckerman, 2007). Further, risk propensity has been associated with negative health outcomes such as pathological gambling (Mishra & Lalumière, 2010). On the other hand, accuracy in personality judgment has been correlated with positive outcomes such as interpersonal control and life satisfaction (Letzring, 2014). Given the importance of making accurate personality judgments and using those judgments to infer risk propensity in others, it is important to understand the accuracy of these judgements. The present study therefore investigated the accuracy of judgments of risk attitudes and propensity in others. The study also investigated the effect of potential moderators (i.e. information relevance, good judge) on the accuracy of these judgments.

Chapter II: Literature Review

Risk

Risk refers to a decision in which one of the options contains a probabilistic or unknown outcome. In line with Sitkin and Pablo (1992), risk is defined within the current project as "the extent to which there is uncertainty about whether potentially significant and/or disappointing outcomes of decisions will be realized" (p. 10). This definition implies several elements of risk that require explication including outcome variability, expectation, and potential, as well as risk behavior. Outcome variability is an aspect of risk that refers to the probabilistic or unknown results in the choice set. Further, variability includes the notion of controllability of those outcomes (Vlek & Stallen, 1980). Specifically, variability refers to estimation by a person of the likelihood of an outcome and their perceived controllability of that likelihood.

For example, Libby and Fishburn (1977) demonstrated that risk decisions made in a business setting were evaluated for risk and preference based on the estimation of outcome distributions. Outcome expectation refers simply to the perceived gain or loss of an outcome (Sitkin & Pablo, 1992). Gains and losses refer to the arithmetic gain or loss of the outcome at hand and does not carry any connotations of goodness or badness of the outcome. Several research programs have established differences in risk-taking behavior based on perceived gains or losses in a variety of areas including choice frames (Kahneman & Tversky, 1979) and game theory (Raub & Snijders, 1997). Outcome potential is the perceived magnitude of a choice outcome or, stated another way, the absolute value of the gain or loss associated with a choice. Outcome potential is associated with overestimations of outcome variability. For example, Allman (1985)

found that individuals were willing to purchase lottery tickets despite the minuscule chance of winning, presumably because the magnitude of the outcome was so great.

Each of the three elements described above leads to different types of risk-taking behavior. Risk behavior is defined as decision making in a risk context. These differences in risk behavior are conceptualized as being a function of the three elements described above. Consider that not all individuals correctly estimate the likelihood of an outcome (outcome variability; Slovic, 2000) or that some individuals have different decision goals and thus, weigh gains and losses differently (outcome expectations; Kahneman & Tversky, 1984). Collectively, outcome variability, expectation, and potential are referred to as risk perceptions (Sitkin & Pablo, 1992; Sitkin & Weingart, 1995). In fact, risk perceptions have been described as the "decision maker's assessment of the risk inherent in a situation" (Sitkin & Pablo, 1992, p. 12). Separate but related to risk perceptions is risk propensity. Risk propensity is a measure of the willingness of a person to embrace risk. This willingness has been measured using both self-report (Blais & Weber, 2006) and behavioral outcomes (Xue et al., 2009). Blais and Weber (2006) measured risk propensity by asking respondents to rate the likelihood they would engage in a particular risk behavior (bungee jumping), while Xue et al. (2009) measured risk propensity by observing actual risky choices in lab settings. Specifically, they asked participants to make several decisions for a small monetary ($.25-$.50) gain or loss. One of the choices yielded a certain outcome while the other involved some level of probabilistic risk. On each successive trial, the probability and outcome magnitudes were varied and the number of times a risky option was selected represented an individual's risk propensity.

Risk perceptions and propensity are considered determinants of risk behavior. As seen in Figure 1, increases in willingness to accept risk and reduced risk perceptions can increase risk behaviors. A higher degree of risk propensity is assumed to lead to more comfort with risk, lower perceived risk, and thus more risk-taking. If people perceive one of their choices as having a low risk and high outcome magnitude, then the likelihood of selecting that option increases. The two determinants also interact to produce varying patterns of risk behavior. For example, consider the person who assesses an option as highly risky with a large outcome but who also has a very low-risk propensity. The lower willingness to take risks reduces the likelihood of selecting the risky option despite a high payoff. In contrast, consider the same high risk, high reward perceptions with someone who has a high propensity for risk-taking, in which case the estimated likelihood of selecting the risky option increases. Empirical evidence supports the determinant model and establishes a need for measuring risk perceptions and propensity with risk behavior (i.e., risk history, decision domain familiarity, and organizational trust; Mayer, Davis, & Schoorman, 1995; Sitkin & Weingart, 1995), while addressing questions surrounding the trait-like nature of these constructs (Blais & Weber, 2006).

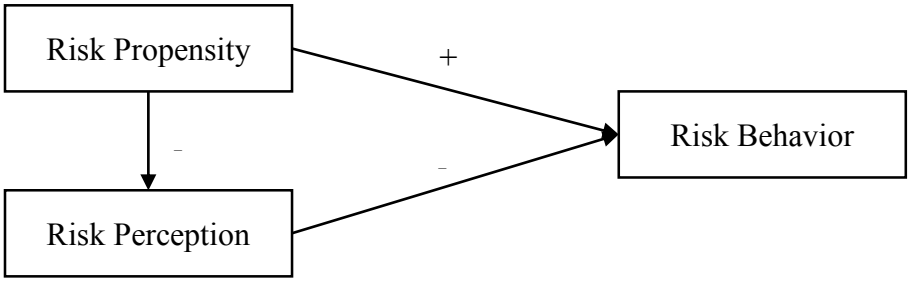

Figure 1. Model of determinants of risky decision-making behavior (Sitkin & Weingart, 1995).

Trait nature of risk. Personality traits refer to the relatively stable and enduring patterns of thoughts, feelings, and behaviors over time (Allport & Allport, 1921; Mischel & Shoda, 1995). In general, conceptualizations and measurement of personality traits have historically been fraught with inconsistencies. Early on, personality traits were thought to be excellent predictors of behaviors, but as researchers collected empirical evidence it became apparent that the situation and environment played a significant role in determining behavior (for a review see Funder, 2009). The finding gave rise to more complex conceptualizations that included the interaction of personality and the environment (Crocker, 2011; Funder & Fast, 2010; Leikas, Lönnqvist, & Verkasalo, 2012; Mischel, 1964, 2004). The debate about whether the situation or personality is more important is considered by many to be closed and replaced by investigating how traits interact with the environment to predict behavior (Fleeson & Noftle, 2009; Funder & Fast, 2010). In line with the investigation of personality traits in general, risk research is moving to consider individual differences in risk propensity and attitudes.

The field of judgement and decision making (JDM) encompasses how individuals combine their desires with their expectations or beliefs to decide a course of action. Research in this area breaks down a decision into a course of action, beliefs about the world, and expected outcomes in order to understand the decision making process (Hastie, 2001). There has been a long-standing reluctance to consider individual differences in JDM in general, despite a shift in all fields to an interdisciplinary approach that includes individual differences (Appelt, Milch, Handgraaf, & Weber, 2011). This reluctance stems from a significant increase in publications using individual differences as mediators and moderators in decision making research. Some individuals in the field

view the practice of using individual differences as mediator/moderators as inappropriate based on issues of interpretability of individual differences and agreement on operationalization (Appelt, Milch, Handgraaf, & Weber, 2011). In a review of the literature in 2001, the use of individual differences in JDM research was described as essentially nonexistent (Highhouse, 2001). From a construct perspective, however, it has been agreed upon that JDM in general includes three elements: decision characteristics (order effects, framing), situational factors (location, constraints), and person characteristics (personality, affect; Einhorn, 1970; Hunt, Krzystofiak, Meindl, & Yousry, 1989). Regardless of the dominant focus on decision characteristics and situational factors, there has been a major shift towards the role of individual differences. For example, researchers in JDM have considered individual differences including motivation (Barrick, Mount, & Judge, 2001), job satisfaction (Judge, Heller, & Mount, 2002), leadership (Judge, Bono, Ilies, & Gerhardt, 2002), and organization commitment (Erdheim, Wang, & Zickar, 2006).

Many of these studies support differences in the relations between decision making and its antecedents at different levels of individual traits. However, measuring risk propensity and perceptions as a trait has yielded inconsistent results. Slovic (1964) found that using different measurement methods yielded different results for risk attitudes. For example, scores on self-report measures of risk propensity differed significantly from behavioral measures like dyadic choice paradigms. Additionally, even when the measurements were the same, people did not show consistency in risk aversion across time and situations (Schoemaker, 1990). MacCrimmon and Weinburg (1986) showed that managers consistently demonstrated one pattern of risk attitude in their

personal life but a different one when making decisions for the company. These inconsistencies in measuring risk attitude as a trait yield poor predictive validity and support the reluctance to consider risk attitudes as a stable trait. One counter-perspective that emerged from the literature is the idea of risk propensity as a "changeable trait that is persistent" (Sitkin & Weingart, 1995, p. 1575). The essential idea is that risk propensity is trait-like (i.e., enduring) but can be modified by learning and experience, which is true of traits in general. Blais and Weber (2006) further proposed conceptualizing risk by its constituent domains as a way to address the trait structure and measurement of risk.

Domain specificity. Given the complex nature of risk behavior, it is also important to consider that not all domains of risk are considered equal by all people. One person may love to bungee jump but never risk a dime of their money, while someone else would take a pill from a stranger at a party but not adopt a social risk by introducing themselves to a new person at the same party. Building on the complex models of person-by-situation interactions, researchers have recently suggested that risk attitudes can be conceptualized in a risk-return framework (RF; Bell, 1995; Jia & Dyer, 1997; Sarin & M. Weber, 1993). The various models of RF treat a risk behavior as an outcome that changes based on the person and the context. In line with a mediated model of determinants of risk decision making (Sitkin & Pablo, 1992; Sitkin & Weingart, 1995), a RF posits that risk behavior is a function of the perceived risk and benefit of a choice a person has in a given situation. The cost benefit process is referred to as risk attitudes. The risk behavior variable itself is broken down into a composition of 1) an evaluation of the perceived risks and benefits, and 2) a trade-off a person makes between units of risk and benefits (Weber & Hsee, 1998; Weber & Milliman, 1997). The evaluation and trade-off are

assumed to be stable across time and situations. For example, Weber and Hsee (1998) found cross-cultural differences in risk perception but cross-cultural similarities in risk attitude when accounting for risk perceptions. The RF and the supporting evidence provide a way to understand how risk perceptions and attitude are dispositional in nature and change as a function of context.

Empirical evidence from the measurement of dispositional risk-taking, attitudes, and perceptions suggests that risk is best understood as being domain specific. Early self-report measurements included several questions from different content areas and then summed the scores for a total risk attitude or perception score (Kogan & Wallach, 1964). For example, the Choice Dilemma Questionnaire asks people to rate their risk aversion to a monetary gamble (financial domain) and participating in a dangerous sport (recreational domain). The problem with this format and scoring is that it gives inconsistent results that can be inflated by one domain. For example, if a person is a lover of extreme sports and rates those items very highly but is a moderate financial risk taker, the recreational items would draw the overall score higher. The subsequent interpretation of the combined score is not easily interpretable due to inflation of the value.

Additionally, there is mounting evidence that people respond differently to content domains based on their affective experience. Affect plays a role in risk perception, with stronger affective responses leading to higher benefit or risk perception during evaluation. The risk-as-feelings theoretical perspective posits that either positive of negative affect experienced during decision making influences that decision. In this framework, the feelings experienced by a presence of risk can and do diverge from the cognitive assessment of that decision (Loewenstein, Weber, Hsee, & Welch, 2001). For

example, when asked to estimate the risk of dreadful events such as a nuclear power plant accident, individuals significantly overestimated the potential risk (Slovic, Finucane, Peters, & MacGregor, 2004). Conversely, positive affective experience impacts decision making as well. Participants who were experimentally manipulated to experience a positive affective state had higher optimism about winning a lottery but were less likely to take a gamble than controls (Isen & Patrick, 1983). Taken together, the evidence suggests decision making involving risk is complex and best understood as domain specific with multiple components affecting the estimation of risk. Therefore, the measurement of risk must take these aspects into account and address risk attitude and propensity alike.

In summary, theories like RF and risk-as-feelings provide evidence of the compound dispositional and domain-specific nature of risk propensity and perceptions. While research is still investigating the nature of these individual differences, there is strong evidence to support the assumption that some aspects of these behaviors are trait-like. The RF and risk-as-feelings perspective provide a conceptualization of risk attitude and perception as a complex interaction of person and situation. Lastly, risk behavior, propensity, and perceptions or attitudes are best understood as domain specific when measuring and assessing the construct. Not all risk is considered equal to all individuals, and one person may find riding a motorcycle without a helmet to have modest benefit and high risk while another might perceive the opposite. In this way, a person is risk-seeking in one domain and risk-averse in the other.

Given the problems with measuring risk across different domains, the Domain-Specific-Risk-Taking (DOSPERT) Scale was designed to better measure the risk attitude

construct (Weber, Blais, Betz, 2002). The DOSPERT allows for the measurement of conventional risk propensity (willingness to engage in domain specific risk) and risk attitudes (perceived risk and benefit of domain specific risk; Blais & Weber, 2006). The proposed experiment seeks to investigate the accuracy of judgment of risk attitude and propensity across these domains, including the role of available information and judge individual differences in risk propensity on these judgments. To accomplish this, we discuss a framework and method for investigating the accuracy of judgment of risk in others given zero acquaintance.

Accuracy of Personality Judgments

Investigating the accuracy of personality judgments began in the first half of the last century with the search for the good judge (Allport & Allport, 1921; Estes, 1938). Essentially, the question was 'what makes a person accurate at judging the personality of another?' As Funder (1995) points out, the initial enthusiasm for answering this question was quelled by three distinct factors that led to a decline of interest over the years. First, initial data collection yielded very low correlations between a judge's ability and personality characteristics. Specifically, correlations between judge's accuracy ability across situations were at the time considered low (Cline & Richards, 1960; Crow & Hammond, 1957). In one study, judge's ratings were taken at three time points over the school year and these ratings showed high item set correlations but a low relationship between accuracy scores (Crow & Hammond, 1957). Second, Cronbach's (1955) critical analysis of the methods used to answer the accuracy question, and his subsequent call for more complex methodology and analysis, unintentionally convinced researchers to abandon the paradigm. Lastly, the introduction of the person perception paradigm

refocused the study of personality judgment on environmental and cognitive factors. Contrived laboratory manipulations, such as lists of trait words, were used as stimuli for judgments, instead of actual people.

The importance of understanding accuracy of personality judgment and the good judge question have regained importance and focus in contemporary research. Chief among models for investigating questions of accuracy is the Realistic Accuracy Model (RAM). The RAM describes accuracy as the degree to which a judgment matches a set of criterion meant to approximate reality (Funder, 1995). While philosophical issues related to absolute truth have been raised in the accuracy literature, Funder (1995) argues that accuracy is a testable hypothesis and meets the standard for scientific inquiry, including the investigation of reliability and validity. Furthermore, accuracy has predictive power than can be scrutinized using appropriate methods. He suggests three standards by which to evaluate the accuracy of personality judgment: self-other agreement, consensus, and behavioral prediction.

Self-other agreement is the degree to which the target and judge agree on the level or nature of a trait. Consensus is the extent to which different judges all agree upon the level of a trait(s) for a target. Funder (2012) describes these first two standards as providing confidence in the accuracy of judgment. If the target and an observer do not agree on a trait level, then someone has to be incorrect. Kolar, Funder, and Colvin (1996) explored using self and acquaintance ratings as a specific criterion to determine accuracy. Targets rated their own personality and acquaintances completed ratings of the target. Results demonstrated that an aggregate of the self and two acquaintance ratings yielded significantly better predictive validity for behavioral criteria than only self-ratings. When

the aggregate is comprised of judges, but not the self, then it is referred to as consensus. Consensus is a piece of accuracy but not a complete measure of accuracy itself. For example, if several observers do not agree on a target trait level, then the accuracy of any one of those judgments is in question. However, a group of judges can all agree on a trait level but still be wrong about the actual level of that trait. Therefore, it should be noted that strong consensus does not equal high accuracy. Finally, behavioral prediction provides the "gold standard" by which all accuracy is evaluated. If a judgment predicts actual behavior in the lab or a more natural setting, then it is said to be accurate.

The process by which accuracy is achieved occurs through the stages of cue relevance, availability, detection, and utilization. Relevance refers to the fact that a cue must be associated with the trait that is being judged. Availability deals with whether or not a cue is present to be detected. Detection involves the process of the judge perceiving the cues, and utilization is the use of the detected cue to make a judgment about the trait. The RAM is an intellectual descendant of the lens model (Brunswick, 1952), and like its predecessor, assumes that relevance and availability occur in the environment, while detection and utilization occur in the perceiver (Funder, 1995). The RAM has been used extensively to organize experiments, and to understand better the accuracy of personality judgment (Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011; Letzring, 2008; Letzring, Wells, & Funder, 2006; Lorenzo, Biesanz, & Human, 2010).

Realistic Accuracy Model

In the description of the RAM, Funder (1995, 2012) points out that the relationship between each of the four elements is multiplicative and can be represented by the following formula:

Accuracy = [(relevance)(availability)] X [(detection)(utilization)].

In the above formula, the groupings of relevance and availability along with detection and utilization exists because the first are considered characteristics of the target while the second two are characteristics of the judge. The implication of the formula is that if any one of the elements is at zero, then accuracy is zero. Imagine if the cue had nothing to do with the trait being judged or the judge never detected the cue. In this case, no accurate impression could be formed. Also, the multiplicity of the model suggests that perfect accuracy is not possible. In a hypothetical example used by Funder, if each of the terms were high at .90, then the overall accuracy score would be only .66. Lastly, the model implies that a significant change in any one of the terms can significantly change overall accuracy. As a result, it has been recommended that research into accuracy investigate and consider each of these stages. Funder also elucidated four key variables that can moderate all of the terms of the model: good judge, good target, good trait, and good information.

Good judge. For accuracy of personality judgment to be achieved, the perceiver or judge needs to detect and utilize relevant and available cues. To this end, people differ in their ability to accurately judge personality. Differences in accuracy based on individual differences in judges have been demonstrated in several experimental paradigms (Christiansen et al., 2005; Letzring, 2008; Letzring, 2010; Powell & Goffin, 2009). Funder (1995) described three judge characteristics that potentially impact accuracy outcomes: knowledge, ability, and motivation. For a judge to have high accuracy, it is theorized that there is a connection between understanding how personality is related to behavior and accuracy of judgments. Powell and Goffin (2009) varied

knowledge by providing training on personality to experimental groups and then measuring accuracy. The results showed increases in accuracy for those in the training groups on several traits. Ability refers to the valid use of cues to infer accurate judgments and is thought to be associated with constructs like general intelligence (Funder, 1995). Christiansen et al. (2005) found that measures of dispositional intelligence and general mental ability predicted the accuracy of personality judgments. Motivation refers to whether the judge cares about making the judgment or whether the context is relevant. Empirical evidence has demonstrated that when individuals are told their accurate judgments will lead to important social outcomes, accuracy is increased (Flink & Park, 1991). Taken together, the judge characteristics of knowledge, ability, and motivation significantly predict increases in accuracy

Additionally, good judges can have behavioral characteristics that elicit better cues from their targets. For example, judges with excellent social skills may elicit more trust and comfort with the target, thereby leading to more and higher quality cues from the target. In fact, Letzring (2008) had previously unacquainted triads interact in an unstructured way and found that agreeableness, social skill, and adjustment of the judge were all positively related to accuracy.

Good target. Some individuals lend themselves well to being accurately judged. Good targets broadly tend to be those with good psychological adjustment, social status, and socialization. Specifically, good targets are those individuals who are open and honest, and whose personalities are easy to understand (Human & Biesanz, 2013). Additionally, good targets tend to be those who are transparent and consistent with their patterns of behavior from one setting to the next, which provides an opportunity for cue

detection and accurate judgment (Funder, 2012). Funder (1995) also points out that individuals who have high trait levels of agreeableness, conscientiousness, and extraversion tend to be judged more accurately. As further support, Colvin (1993) found that expressive and extroverted people are more accurately judged than their peers. Overall, the idea is that a good target is one who provides multiple, consistent, higher quality cues, and who tends to be well-adjusted with good social skills (Funder, 2012).

Good trait. The RAM model suggests that traits that are easier to observe are easier to detect, and thus are more accurately judged (Funder, 2012). Consider the example of the extravert who is high on sociability and regularly talks in social situations. In this case, extraversion is easy to observe, and judgment can proceed in a manner that is likely to result in high accuracy. In contrast, consider the person low in emotional stability. His frequent worrying or emotional sensitivity might occur mostly on the inside and thus is not very observable to others. In this way, some traits are easier to observe and easier to judge accurately. There are caveats to this moderator. First, Vazire (2010) developed the Self-Other Knowledge Asymmetry (SOKA) model to understand the relationship between the visibility of traits and accuracy of judgment. Some traits, like emotional stability, often manifest privately and have low observability, and are better judged by the self; while traits high in social desirability are best judged by others. In addition to these findings, John and Robins (1993) found that desirability of a trait is related to accuracy. People tend to be motivated to hide traits that are considered undesirable and display traits that are desirable in order to maintain a positive sense of self. For example, someone who is low in conscientiousness and demonstrates low reliability may take pains to hide that fact from someone they have just met. The cloaking

of low desirable traits makes detection and accurate judgment more difficult. Traits that are high in desirability, however, lend themselves well to visibility and thus, accurate judgment.

Good information. The quality and quantity of information available for detection provide two possible paths to accuracy (Funder, 2012). In a simple test of this moderator, Blackman and Funder (1998) varied the amount of time for a judge to observe the target (5 or 30 minutes) and found that longer observations significantly improved accuracy. Biesanz, West, and Millevio (2007) also demonstrated that acquaintanceship is related to accuracy. Using a model of accuracy that includes stereotypical and differential types of accuracy, they showed that the length of time someone is known impacts these two type of accuracy. They found that as length of acquaintanceship goes up, differential accuracy increased and stereotype accuracy decreased. In comparison, information can add to accuracy by providing relevant quality cues for utilization. Imagine a person who talks specifically about their personality for 5 minutes versus someone who talks for the same length of time about their favorite hobbies. In fact, Letzring, Wells, and Funder (2006) found that the relevance of information to personality significantly increased accuracy. The accuracy of personality judgment is significantly moderated by the quality and quantity of information (Funder, 2012) and is an important part of understanding the paradigm.

The study of what types of information are of high quality and relevant to accuracy is fairly new (Letzring, Wells, & Funder, 2006; Letzring & Human, 2013). The RAM model has been utilized since 1995 to model accuracy, but a significant gap in the research on what types of information are important remains. Consider a simple scenario

in which a person gains information about a target's thoughts and feelings in comparison to observing their behaviors across situations. Is there something distinct about each of these types of information that leads to different levels of accuracy in personality judgment? In a study of information quality, Letzring and Human (2013) examined how accuracy changed as a function of whether judges watched a video where targets talked about their thoughts and feelings, talked about behaviors across several situations (e.g., with family or at work), or performed various behavioral tasks (e.g., playing Jenga). Judges then rated each of the targets on the Big Five personality domains. The authors used the RAM, and a separate model of accuracy developed to understand the components of accuracy, the Social Accuracy Model (SAM; Biesanz, 2010). The SAM is a statistical model of accuracy that partials out normative accuracy – the judge's ability to perceive others in a way that is consistent with what the average person is like on a trait or dimension – from distinctive accuracy – judging how a target is above, below, or at the mean for that trait. The RAM posits that individual differences within each judge can and do influence the detection and utilization stages of accuracy. The RAM was used to conceptualize the stages of accuracy and the good information moderator while the SAM was used to separately examine normative and distinctive accuracy. Results from this study demonstrated an important aspect of information quality. Judges were higher in distinctive accuracy than normative accuracy after observing the thoughts and feelings and behavioral discussions for some traits, compared to observations of a target's behavior. The conclusion from this study, and others on information relevance, demonstrates that accuracy does change as a function of information quality. Going

forward, more research is needed to understand the moderating effect of information relevance.

Realistic Accuracy Model and Risk

A large body of literature demonstrates that accuracy of personality judgment is relevant and can be investigated. The RAM model is an extremely useful model for organizing and conducting research on the accuracy of personality judgments. Specifically, the RAM is functional for generating testable hypotheses. While there has been a sizable body of research looking into accuracy or personality judgment for everything from the five-factor model to behavioral outcomes, to date there has been a minimal inquiry into the accuracy of judgment of risk propensity and attitudes. Researchers have examined the ability of people to infer risky behavioral outcomes from faces (Carré, McCormick, & Mondloch, 2009; Olivola & Todorov, 2010; Stillman, Maner, & Baumeister, 2010; Valla, Ceci, Williams, 2011). Stillman et al. (2010) asked whether actual violent tendencies could be accurately estimated after a brief exposure to photographs of registered sex offenders. Results indicated that actual violent acts estimated by participants correlated highly with the arrest record of the sex offenders. Valla, Ceci, and Williams (2011) gave participants either pictures of convicted criminals or noncriminal faces. Participants were able to accurately infer criminality from the photographs alone after controlling for age, race, and gender. Carré, McCormick, and Mondloch (2009) showed that aggressive propensity could be accurately judged when participants were exposed to a photograph of only a person's face.

While the research elucidated here lays the groundwork for a possibility that risk behaviors can be accurately predicted, there are several problems. First, each of these

studies uses photographs only and, based on literature, there is a strong possibility that the amount and quality of information provided in a live interaction of video would yield different results. For example, Borkenau & Liebler (1992) video recorded targets entering a room, sitting down, looking into the camera, and speaking. These video clips were edited into audio only, silent film, stills and full video clips and shown to a sample of judges. Results demonstrated that varying the types of information via clips yielded differences in the validity but not reliability of judge rating of targets on personality. Of note, judges watching the video with sound demonstrated the highest level of accuracy across personality traits with the exception of conscientiousness. Further, Naumann et al. (2009) showed that even when using photographs, different information leads to differences in personality judgements. Judges who viewed targets who were asked to make a spontaneous pose for a photograph were more accurate across all traits being judged than judges who viewed photos of targets with an emotionless and constrained pose. Second, the behaviors discussed here are more problematic and extreme, such as violence and aggression. It would be useful to know if more common risk-taking behaviors and attitudes (e.g., recreational, financial) can be inferred from a normal population of targets. Third, these studies do not provide a direct test as to how accurately risk attitudes and propensity can be judged.

One study has looked specifically at the accuracy of risk attitudes, perception, and behaviors. Mishra and Sirithran (2012) examined whether first impressions derived from pictures would lead to accurate judgments of risk. Participants were given standard self-report measures or risk propensity (i.e., DOSPERT, Zuckerman's Sensation-Seeking Scale) along with behavioral measures of delay discounting and gambling. Then a photo

was taken of each person, and a second set of participants viewed each of these pictures and were asked to rate each picture by answering the question "how risky do you think this person is?" from 1 (*not at all risky*) to 10 (*very risky*). The results indicated that when a pooled average of the risk ratings was compared to the self-report and behavioral measures, judgments were positively correlated with scores on each measure. This result provides evidence that risk propensity and behaviors can be accurately estimated; however, it used pictures rather than more informative stimuli such as video. Also, the authors did not examine any of the moderators of accuracy identified by Funder (1995).

An additional limitation of the existing research in this area is that the studies used general and one-dimensional measures of the risk behavior or propensity that do not take into account the multidimensional nature of risk. In the Mishra and Sirithran (2012) study, participants were asked to rate risk using a single question. The main issue with this is that risk has been shown to be multidimensional, and as previously discussed, risk-taking behaviors and attitudes in one dimension can artificially inflate or deflate overall scores (Blais & Weber, 2006). The authors administered the DOSPERT to the targets and correlated those scores with the single risk rating of the judges. This method leaves significant information unaccounted for in contrast to having judges rate a target for each of the DOSPERT items. Furthermore, well-established and validated methods for assessing the construct of accuracy (i.e., RAM) suggest that the issue of accuracy is complex. The complexity means that the moderators of accuracy should be considered when modeling accuracy. In addition, measures of risk need to be evaluated for validity and reliability using the standards of assessing accuracy elucidated in the RAM. As

Funder (1995) demonstrated, a correlation between a single judgment and a measure of personality is not sufficient to evaluate accuracy.

Hypotheses

Thus, the first hypothesis of the current experiment was that individuals would be able to accurately judge trait risk-taking (i.e., attitudes and propensity) in others. Research using still photographs strongly suggests that accurate perceptions are possible with as little information as is available in a still picture. A considerable body of literature on the accuracy of traits has further established both normative and distinctive accuracy are possible across different traits. However, this work has not yet been extended to trait risk propensity or attitudes. Further, the current experiment was interested primarily in distinctive accuracy compared to normative accuracy. Distinctive accuracy refers to the accuracy of judging targets as similar to or different from other targets and the average person. The focus is on distinctive accuracy because we are interested in how accurate people are at judging how targets are different from the average person.

The second hypothesis pertains to the moderator of information quality. The research conducted so far suggests that both information quantity and quality differentially affect accuracy, but no studies to date have examined this moderator in judgments of trait-level risk. Results from different experiments suggest that as information quality increases, so too does distinctive accuracy. Letzring and Human (2013) demonstrated that watching different targets discuss thoughts and feeling compared to discussing behaviors across situations leads to a different effect on accuracy. It was therefore hypothesized that as information quality increased, judge's distinctive accuracy would also increase (Figure 2). Specifically, judges would demonstrate the least

amount of accuracy in the low relevance condition and the most accuracy in the specific relevance condition. Further, judges in the broad relevance condition would show higher accuracy than the low relevance but less accuracy than the specific relevance conditions.

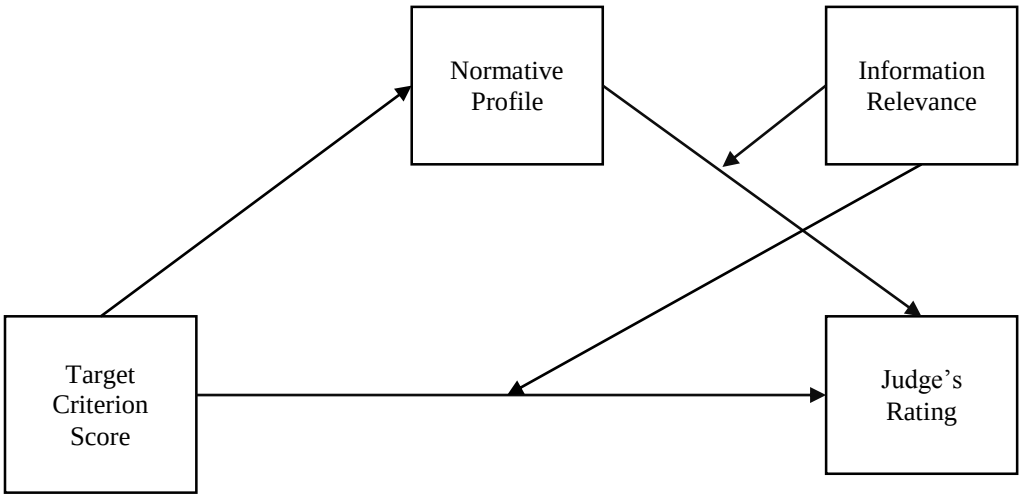

Figure 2. Model of hypothesis 2 showing moderator relationship of information relevance on normative and distinctive accuracy paths.

The third hypothesis concerns the interaction of the good judge and good information moderators, or what Funder (1995) refers to as sensitivity. Empirical evidence demonstrates that characteristics of the judge and information quality are related to accuracy. For example, as judgemental ability or intelligence scores increase, so does accuracy. Currently, research into the sensitivity interaction lacks empirical findings, and it is the intention of the current experiment to contribute to that literature. Based on combining research from both moderators, it was hypothesized that the individual difference of risk-taking in the judges would interact with judge-target ratings and information relevance to produce higher accuracy in low-quality information circumstances when judge risk propensity is high (Figure 3).

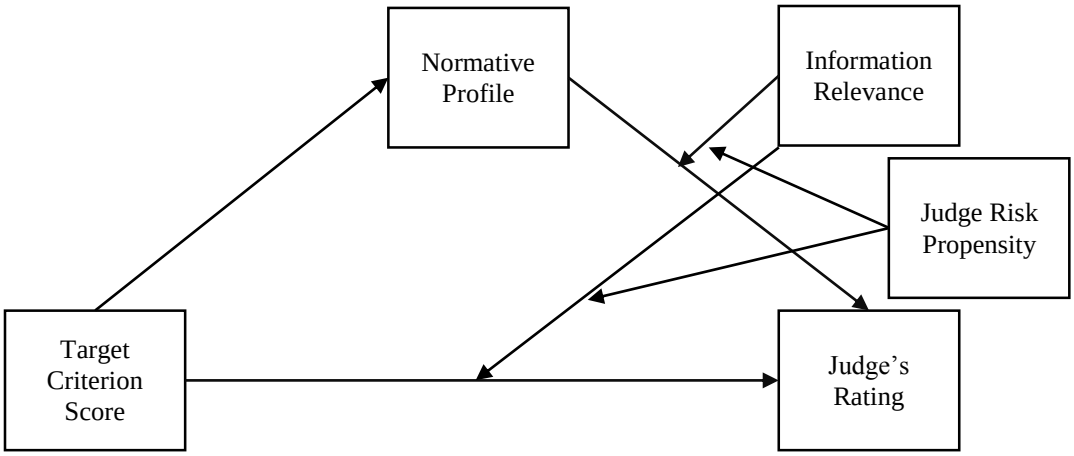

Figure 3. Model of hypothesis 3 showing three-way interaction of judge risk propensity and information relevance on the normative and distinctive paths.

Present Study

This paper proposes to replicate previous findings in the accuracy of personality judgment literature by using the RAM and SAM to investigate the accuracy of judgments about risk propensity and attitudes. Specifically, accuracy will be assessed using the RAM in that the effect of good information on accuracy will be experimentally examined. The SAM will be used as a computational model to partial out the individual components of normative and distinctive accuracy. An accuracy criterion will be established by using an aggregate of self and acquaintance ratings. The effect of good information will be assessed by creating three experimental manipulations of information quality. With previous work on information relevance in mind (Letzring & Human, 2013), relevance will be manipulated to have three levels – low relevance, broad relevance, and specific relevance – which will be used to address the impact of information relevance on the accuracy of judging risk in others. The low relevance condition will contain only information about broad personality traits (e.g.,

agreeableness) and no information about target risk-taking. The broad relevance condition is so named because targets provided verbal information about their attitudes towards risk in general (e.g., risk as having a positive or negative valence). Finally, the specific relevance condition contained a specific behavioral example of risk-taking from the target's life.

The current experiment further seeks to broadly replicate the finding of Mishra and Sirithran (2012) by examining the question of whether risk attitudes and propensity can be accurately judged by others using new methods (videos as opposed to pictures), multidimensional measurements of risk (i.e., DOSPERT), and empirically established models of accuracy (i.e., RAM, SAM). The proposed experiment will further extend the previous work by examining some of the moderators of accuracy (good judge, good information).

Based on the above information, the following hypotheses were generated:

H_1: Overall, risk attitudes and propensity will be accurately judged across targets.

H_2: Information relevance will affect the accuracy of risk judgments. Judges in the specific relevance condition will be more distinctively accurate than those in the low relevance and broad relevance conditions.

H_3: Individual differences in judge risk propensity will moderate accuracy of judgment. Specifically, higher risk propensity will increase accuracy more in the low relevant condition than the broad relevance and specific relevance conditions, whereas lower risk propensity will have the same effect on accuracy across all conditions.

Chapter III: Methodology

Method

Overview

An experimental, multi-level model design was used to investigate the hypotheses. To accomplish these goals, the current experiment replicated the methods of previous accuracy experiments (Letzring & Human, 2013). First, stimuli were created by bringing targets and two of their acquaintances who had known them for at least 6 months into the lab to complete a measure of risk (DOSPERT). Targets completed the measures for themselves, and the acquaintances provided ratings of the targets. The target and acquaintance ratings were combined to create an accuracy criterion (Kolar et al., 1996). Specifically, acquaintance ratings of the target for each item of the DOSPERT were averaged and then that aggregation was averaged with the targets rating of themselves. The targets then completed a short interview, during which they described their personality, their attitudes and thoughts about risk, and a risky behavior in which they had engaged. A second set of non-acquainted participants viewed the interviews and provided judgments of the targets, and also completed the same self-report measures.

To investigate the effect of information relevance on accuracy, approximately one-third of the participants viewed the personality segment of the video (low relevance), one third viewed the risk attitudes and thoughts portion of the video (broad relevance), and one-third viewed the section of the video describing a risky activity (specific relevance). Participants were randomly assigned to risk relevance group. The low, broad, and specific relevance conditions are the independent variables in the design, while the

relationship between judge ratings and the accuracy criteria is the dependent variable. The judges' self-reported ratings of risk propensity were examined as a potential moderator of the relationship between information condition and accuracy to investigate the moderating effect of the judge's risk-taking on the accuracy of judgments.

Determination of Sample Size

To determine the sample size for the current experiment, research by Maas and Hox (2005) was taken into account. In their review, they pointed out that group sizes as small as 30 have been used in previous research using multilevel models. Based on simulations using Monte Carlo methods that varied the interclass correlations and group sizes, the authors concluded that a minimum of 100 total groups (judges) provided acceptable regression coefficients and variance estimates. Biesanz (2010) extended this work and demonstrated that including 100 judges total provided acceptable model estimates. Based on the previous research, the sample size needed to produce an interclass correlation of .30 and detect a medium effect was calculated using Optimal Design (Raudenbush, 2011). Based on this analysis, a total sample size of 250 was collected for the current experiment[1]. An *ad hoc* computation demonstrated an acceptable level of interclass correlation for this model (.31).

Participants

A total of 250 participants were recruited from Idaho State University's Psychology Department participant pool (72.16% Female; 70.32% Caucasian, 17.80%

[1] The original proposal recommended 450 participants but was later revised after further background search, power analysis using Optimal Design, and consultation with members of the committee who are experts in multilevel regression models.

Hispanic, 2.40% Black/African American, 2.32% Other; M_{age}=21.43, SD_{age}=5.13, $Range_{age=}$18-53 years). Of the 250 participants' data collected for the current experiment, 4 were found to be missing all of their experimental data due to computer errors early in the experiment. The participants were removed and the interclass correlation was calculated again with a new *n*=246, but this did not change from the previously calculated interclass correlation of .31. Going forward, all analyses were conducted using the remaining 246 participants[2].

Measures

Domain-Specific Risk-Taking. The DOmain-SPecific Risk-Taking (DOSPERT; See Appendix A; Blais, & Weber, 2006) Scale is a self-report instrument that measures the level of risk propensity and attitudes. The DOSPERT is a 30-item measure broken down into five separate content domains: ethical, financial (further decomposed into gambling and investment), health/safety, social, and recreational decisions. Each item represents a specific risky behavior (i.e., drinking heavily at a social function). Each of the 30 items are rated for how likely the respondent is to participate in the activity given the opportunity from 1 (*extremely unlikely*) to 7 (*extremely likely*), how risky the activity is perceived to be from 1 (*not risky at all)* to 7 (*extremely risky*), and how much benefit can be gained from the activity from 1 (*no benefits at all*) to 7 (*great benefits*). In previous research, the internal consistency reliability estimates for DOSPERT risk-taking

[2] Four attention checks were embedded throughout the experiment to address participant fatigue with the tasks and systematic nonresponding patterns. These four checks simply asked the participant to select a certain answer for a self-report question. The plan was to exclude anyone who failed two or more checks. Six participants missed two of the four checks. Analysis of all hypotheses with and without these participants yielded no differences in estimation or interpretation and therefore all data were retained.

scores ranged from .71 to .86, and those associated with the risk-perception scores, from .74 to .83 (Weber, Blais, & Betz, 2002). The current data yielded good reliability coefficients for the risk-taking scale (.78), risk perception scale (.83), and risk-benefit scale (.81). The DOSPERT was used in three ways: as a self-report for participants, as an other report for acquaintances to rate their associated target, and as an other-report for judges to rate targets. The final scores were used to calculate the average rating for each item and used as the normative profile in order to estimate normative accuracy using the SAM.

General demographic questions. General demographics included age, gender, ethnicity, household income, employment, education, GPA, and marital status.[3] Data from income, employment, education, and marital status were used to calculate socioeconomic status using the four factor index (Hollingshead, 1975). All other demographic data were used to describe the sample from the current experiment.

Stimulus Creation

Participants in the main experiment watched seven video clips, each approximately 1 minute in length, of targets being interviewed. To create the stimulus materials, participants were asked to come to the lab and bring two acquaintances who had known them for at least 6 months (*M*=16.53 years, with a range of 3.2-27.2 years).

[3] In addition to the main empirical focus of this paper, four other measures were included for future analysis. The Big Five Inventory (BFI; John, Naumann, & Soto, 2008), Brief Sensation Seeking Scale (BSSS; Hoyle, Stephenson, Palmgreen, Lorch, & Donohew, 2002), CUPS task (Weller, Levin, Shiv, and Bechara, 2007), Delay discounting (Kirby & Maraković, 1996) and a single question of risk judgment (Mishra & Sritharan, 2012), "How risky do you think this person is?"

To reduce the chance that the main participant or target would be familiar with judges, the stimulus material participants were recruited from the Idaho Falls campus of ISU.

Table 1

Target and Judge Demographic and DOSPERT Scores

	Target *M*(*SD*)	Judge *M* (*SD*)	Norms *M*(*SD*)
n	12	246	359
Gender			
Male	3	74	191
Female	9	172	168
Age	24.48 (7.47)	21.43 (5.13)	
GPA	2.93(.77)	3.29(.54)	
SES	35.33(3.96)	37.21(2.24)	
Ethnicity			
Caucasian	75%	78.5%	
Hispanic	25%	17.8%	
African American		2.4%	
Other		2.3%	
DOSPERT			
Financial	13.83(4.99)	12.44(4.69)	19.61(7.73)
Recreational	28.92(6.46)	23.58(8.38)	23.43(9.14)
Social	30.67(4.27)	28.57(4.79)	32.58(5.69)
Health	23.42(8.07)	20.45(7.53)	20.63(7.43)
Ethical	13.83(4.80)	13.21(4.92)	16.92(6.59)

Note. Norms for the DOSPERT were derived from an international sample (*n=359)* reported in a previous study (Weber & Blais, 2006).

A total of 12 participants (see Table 1 for demographics) and their acquaintances were greeted when they came to the lab, seated at a table, and provided with a laptop computer. The acquaintances needed to have known the participant for at least 6 months,

and the researcher verbally verified this before continuing. The study only continued if both acquaintances were present and had known the participant for at least 6 months. The primary investigator[4] asked the participant and his/her acquaintances to read the informed consent and sign and date if they consented to participate. The consent was presented to the participant in paper form and on the computer for the acquaintances. The experimental materials were all presented using MediaLab (Jarvis, 2008).

When the participant and his/her acquaintances indicated they wished to continue, the researcher summarized the experiment and specifically highlighted that the main participant would be videotaped during an interview in which they would be asked to describe their personality in general, their attitudes towards risk-taking, and a risky activity they had engaged in that they were comfortable sharing. Participants were next asked to complete the brief demographic questions and the DOSPERT. Each acquaintance completed the DOSPERT about the participant by answering how they see the target on each item. The participant and acquaintances were then asked to record how long they have known each other and answer basic demographic questions about themselves (e.g., age, gender, ethnicity)[5]. After the acquaintances and participant had completed the measures, the acquaintances were dismissed. Before beginning the interview, the researcher underscored that their video would be shown to other students, and, therefore, they should not describe any risky activity they would not want shared

[4] The primary investigator for this dissertation conducted all interviews for stimulus material creation.
[5] Acquaintances were not presented with items measuring SES.

with others or information that has the potential for negative consequences. The participant was then asked again if he/she wished to continue.

Next, a video camera was used to record a short interview with the participant about his/her risk-taking. Before beginning the interview, the participants were given a piece of paper with the questions they would be asked during the interview (see Appendix B). They were given between 5 and 10 minutes and asked to write a brief reply to each of the questions. The time was alloted to familiarize the participant with the questions before actual video recording, increase comfort to facilitate more natural responses, and allow participants to consider their answer to the risk activities question in advance. After they had completed the answer form, the video recording session commenced. The PI acted as the interviewer for all stimulus creation interviews. The interview began with introductions and simple warm up questions like, “tell me about what you like to do for fun” or “what are you interested in studying?” Next, using a semi-structured format, the interviewer asked the questions from the preset script. From start to finish the interview lasted approximately 5 minutes.

At the end of the session, participants were reminded again that their video interview would be shown to future participants and asked again if they still consented to have their data used. Specifically, they were provided with a written consent form that addressed the use of their video interview and provided a signature line to indicate their consent to use the materials. This consent restated from the main consent form that they may withdraw their consent at any time including at a later date after the session is complete.

After completion, participants were presented with a debriefing form explaining the general nature of the experiment, risk, and providing contact information for the principal investigator and faculty supervisor. Finally, participants were thanked for their participation, excused from the lab, and four SONA credits were granted to the participant. Acquaintances were also thanked again for their time but were not compensated in any way.

From the stimulus creation sample, data and video interviews from seven participants were selected to be used in the main experiment. To create a diverse pool, targets were chosen by examining their individual scores on the DOSPERT and comparing each to the mean and standard deviation of the sample (see Table 1). Also, the sample was compared to a broad cross-section reported in Blais and Weber (2006) to assess this sample's relationship to empirical norms and investigate possible floor effects based on the convenience sample. The comparison shows the target sample is within one standard deviation of the mean from the comparison sample (Table 1). Since all participants yielded complete data and video clips with material appropriate for the main study, all targets were considered.

The final seven were selected by first choosing an individual at the mean and then one standard deviation above and another one standard deviation below the average. The next four were then chosen to even the number of men and women represented in the videos. From the seven targets who were selected, three separate videos were created. One contained the answers to the general personality questions, one for the attitudes towards risk questions, and finally one for the discussion of a risky activity. This yielded

a total of 21 videos to be used in the main experiment (see Appendix C for specific examples of target responses).

Accuracy Criterion

As Funder (1995) has pointed out, there are several ways to construct an accuracy criterion. Kolar et al. (1996) demonstrated that obtaining ratings from acquaintances of the target to aggregate with target self-ratings provided an acceptable accuracy criterion. To compute an accuracy criterion score for the DOSPERT, acquaintance responses for each item were averaged with each other and then that aggregate was averaged with the target's self-report response on that same item (see Figure 4 for a visual of this process).

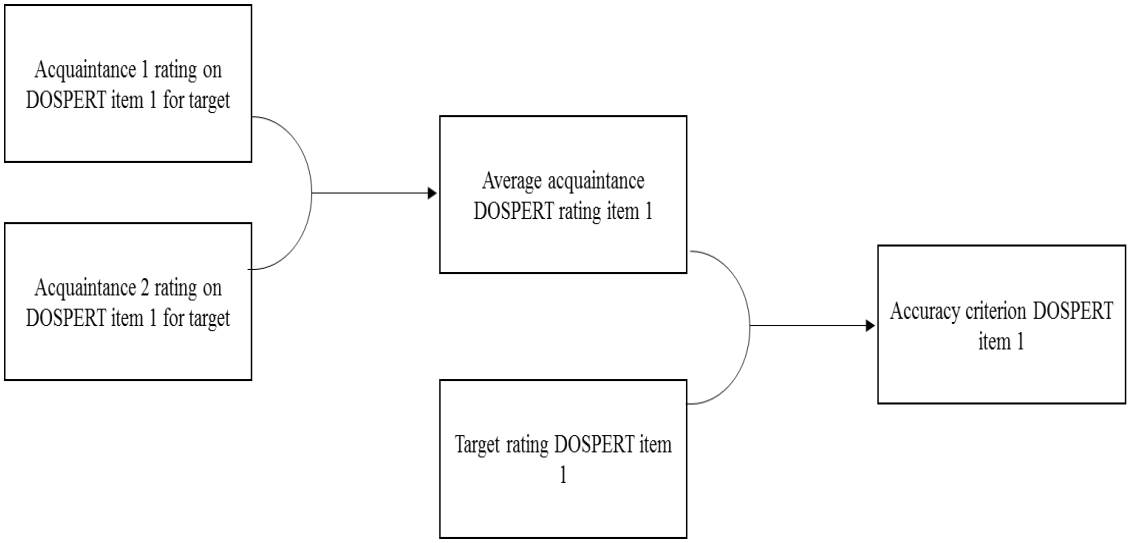

Figure 4. Visual representation of accuracy criterion creation.

Procedure

Participants were first asked to read the consent form and indicate if they wish to continue. The participants were then asked to take a seat and complete questionnaires on the computer using MediaLab (Jarvis, 2008) experimental software. First, participants

completed the 90-item DOSPERT for themselves. Next, the participants were randomly assigned to an experimental condition. All judges observed the same seven targets, but each watched different parts of the interview based on their information relevance condition. Judges in the low relevance condition watched video clips of targets talking about their personality in general, those in the broad relevance condition watched clips of targets discussing their attitudes risk, and those in the specific relevance condition watched clips of targets giving a specific example of the riskiest activity they have ever participated in. Targets were presented in random order to each judge. This was accomplished by programming MediaLab software to generate a random order for the set of targets.

After each video clip, participants were asked to complete the DOSPERT for the target they had just viewed, with directions instructing the judges to estimate the characteristic or answer for the target. For example, instead of asking the judge to rate the likelihood they would engage in "admitting that your tastes are different from those of a friend," they were invited to rate the probability that the target would engage in "admitting their tastes are different from those of a friend." Last, participants completed the demographic questionnaire. Upon completion of the target ratings, the participant was debriefed, thanked for their participation, and excused.

Analytic Approach

Recent quantitative work on the accuracy of personality judgment has yielded a specific analytic approach to deriving the variables needed to study accuracy. The SAM (Biesanz, 2010) is a cross-classified multilevel linear model (MLM) that takes into account both the normative and distinctive aspects of accuracy. MLM is used to analyze

data that vary at more than one level and are nested within clusters. A classic example of nested data used in a MLM is student scores on a standardized test (level 1) being nested within different schools (level 2). The current experiment contains data with judges and targets as two level 2 units. When the data structure contains two level 2 units and level one data are nested within both, this is referred to as a cross-classified model. Judge ratings of each target are nested within individual judges (level 2, unit 1) and within each target (level 2, unit 2). Recall, normative accuracy refers to a judge's ability to perceive others in a way that is consistent with what the average person is like on a trait or dimension. Distinctive accuracy, in contrast, is the ability of a judge to differentiate a specific individual's self-reported trait as above, below, or at the mean of the normative self-report profile on that trait (Biesanz, 2010).

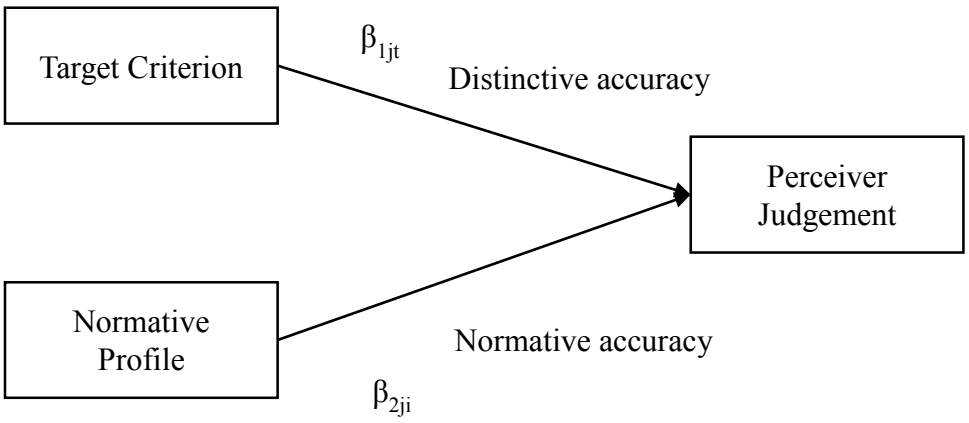

Figure 5. A theoretical model of the accuracy of personality judgment, Social Accuracy Model (Biesanz, 2010).

Thus, the SAM was used to test whether judges, regardless of condition, could achieve a statistically significant level of distinctive accuracy for judgments of risk in others (H_1; Figure 5). Further, the SAM was utilized to investigate the moderating effect

of information relevance (H_2) on judge-target accuracy and judge risk-taking (H_3) as a moderator of information relevance and accuracy.[6]

The following Equation (1.1) represents the components of the SAM used to estimate judge ratings:

$$Y_{jti}=\beta_{0jti} + \beta_{1jt}Tdist_{ti} + \beta_{2jt}Norm_i + \varepsilon_{jti} \quad (1.1)$$

$$\beta_{0jt}=\beta_{00} + u_{0j} + u_{0t} + u_{0(jt)} \quad (1.2)$$

$$\beta_{1jt}=\beta_{10} + u_{1j} + u_{1t} + u_{1(jt)}$$

$$\beta_{2jt}=\beta_{20} + u_{2j} + u_{2t} + u_{2(jt)}$$

The level 1 equation (1.1) represents each judge's ratings of the individual targets while the level 2 equations (1.2) represent the judge-target dyads. Using this model, Y_{jti} represents judge *j*'s rating of target *t* on item *i* of the DOSPERT. $Tdist_{ti}$ represents each target's accuracy criterion on item *i*. $Norm_i$ is the mean self-report on item *i* taken from the overall judge sample. $Norm_i$ was calculated by averaging all judge scores (*n*=246) for each item of the DOSPERT to represent the normative profile for each of those items. For each judge-target dyad, the estimated regression coefficient β_{0jt} is the intercept. The estimated coefficient β_{1jt} is an estimate of judge-target accuracy for judge *j* with target *t* after holding $Norm_i$ constant. Explicitly, β_{1jt} is the estimated level of distinctive accuracy,

[6] The intraclass correlation (ICC) is the amount of variance in the outcome variable accounted for by level 2 cluster membership or the between cluster effect. In order for MLM methods to be appropriate, a sufficient ICC is required because if the ICC is too low it means cluster membership does not contribute to the outcome variable enough to justify a MLM and a more parsimonious statistical analysis is recommended. Therefore, an ICC with a conventional cutoff of .30 was calculated on the following models to determine if analysis using MLM is appropriate. Results yielded an ICC of .31.

controlling for the average person's self-reported risk. Last, β_{2ij} is the estimated level of accuracy between judge *j* and the normative profile score on item *i* after partialling out target *j*'s self-report on item *i*. Explicitly, β_{2jt} is the estimated level of normative accuracy, controlling for the target's distinctive accuracy criterion for risk.

The fixed effects - $\beta_{00,}$ β_{10}, and β_{20} - represent the average intercept, distinctive accuracy slope, and normative accuracy slope across judges and targets. Each remaining term in Equation 1.2 represents the random effects in relation to the grand mean. For example, u_{1j} is judge *j*'s distinctive accuracy slope averaged across seven targets, u_{1t} is target *t*'s average distinctive accuracy slope averaged across 246 judges, and $u_{1(jt)}$ is the accuracy of judge *j* assessing target *t* after partialling out the grand mean (β_{10}) and main effects of the judge (u_{1j}) and target (u_{1t}). Judges (e.g., u_{1j}), targets (u_{1t}), and dyads (e.g., u_{1jt}) are all allowed to vary from the grand mean. β_{10} is the grand mean of the distinctive accuracy slope, and u_{1j} is the unique, distinctive accuracy slope for judge *j*. Therefore, β_{10} + u_{1j} is judge *j*'s unique, distinctive accuracy averaged across all targets. Table 2 shows the interpretation of the main components of the SAM. It should be noted that before all analyses, the variables were grand mean centered. Grand mean centering was used to reduce computational complexity and improve interpretation of estimates in the model. Grand mean centering has no effect on the normative and distinctive estimates in equations 1.1 and 1.2. Additionally, subtracting the mean normative score on each item of the DOSPERT from each target-specific accuracy criterion from that item creates orthogonal predictors. Thus, the model will be used to evaluate H_1, that regardless of condition, judges will be able to accurately perceive target risk.

Table 2

Social Accuracy Model Components

	Perceiver	Target
Distinctive Accuracy	$\beta_{10} + u_{1j}$ The degree to which one perceives the unique risk traits of others	$\beta_{10} + u_{1t}$ The degree to which one's unique risk traits is perceived by others.
Normative Accuracy	$B_{20} + u_{2j}$ The degree to which one's perceptions of others matches the average person.	$B_{20} + u_{2t}$ The degree to which one is similarly perceived to the average person.

Note. Based on a figure originally presented in previous research (Biesanz, 2010).

To compare distinctive accuracy across information conditions (H_2), a groups code approach was used in which the low relevance and broad relevance conditions were used as reference groups in two separate models. Three dummy code variables were created to identify which condition the participant was in: low relevance (C1; 0=no, 1=yes), broad relevance (C2; 0=no, 1=yes), or specific relevance (C3; 0=no, 1=yes) condition. In the first model, low relevance and broad relevance were entered as moderators of normative and distinctive accuracy, which means the specific relevance condition is the reference group. In the second model, low relevance and specific relevance were entered into the model as moderators of normative and distinctive accuracy and the broad relevance condition served as the reference group. Using both models is necessary to determine the comparison for all conditions to each other. Each of these terms were added to the model at level 2 as moderators of accuracy (Letzring & Human, 2013). The moderator term for condition was entered at level 2 because judges were randomly assigned to each condition and watched different videos of the seven

targets. Therefore, condition moderated both judges and targets. The following Equations (2.1, 2.2) demonstrate the model:

$$Y_{jti}=\beta_{0jt} + \beta_{1jt}Tdist_{ti} + \beta_{2jt}Norm_{k} + \varepsilon_{jti} \quad (2.1)$$

$$\beta_{0jt}=\beta_{00} + \beta_{01} + u_{0j} + u_{0t} + u_{0(jt)} \quad (2.2)$$

$$\beta_{1jt}=\beta_{10} + \beta_{11}C1_{j} + \beta_{12}C2_{j} + u_{1j} + u_{1t} + u_{0(jt)}$$

$$\beta_{2jt}=\beta_{20} + \beta_{21}C1_{j} + \beta_{22}C2_{j} + u_{2j} + u_{2t} + u_{0(jt)}$$

The interpretation of these equations remain the same with the exception of β_{11}, β_{12}, β_{21}, and β_{22}. The new coefficients β_{11} and β_{12} represent the average level of distinctive accuracy change between that condition and the reference group in each analysis. The interpretation also applies to normative accuracy, where β_{21} and β_{22} now represent the average level of normative accuracy change between that condition and the reference group. In the first analysis using dummy coding, low relevance was assigned as the reference group. Therefore, the partial regression coefficients in the output represent the relationship between low relevance and broad relevance and also low relevance and specific relevance conditions. In order to capture all condition comparisons, a second analysis with broad relevance coded as the reference group was conducted. The partial regression coefficients in the second model gave the final comparison of broad and specific relevance conditions.

To test the moderating effect of judge risk-taking on information relevance and accuracy, the continuous variable of judge risk-taking scores from the DOSPERT were entered into the model in the level two equations. First, judge risk propensity scores from the DOSPERT were entered as a continuous moderator of accuracy. This model uses

equations 2.1 and 2.2, except the continuous risk variable is paired with the normative and distinctive components. Since judge risk is continuous, no additional dummy coding was necessary to conduct this analysis. Next, two models were constructed by adding judge risk propensity scores as a moderator to the dummy code for the information relevance condition and both types of accuracy.

$$Y_{jti}=\beta_{0jt} + \beta_{1jt}Tdist_{ti} + \beta_{2jt}Norm_{k} + \varepsilon_{jti} \quad (3.1)$$

$$\beta_{0ij}=\beta_{00} + \beta_{01} + u_{0j} + u_{0t} + u_{0(jt)} \quad (3.2)$$

$$\beta_{1jt}=\beta_{10} + \beta_{11}Jrisk_{j} + \beta_{12}C1_{j} + \beta_{13}C2_{j} + \beta_{14}C1_{j}*Jrisk_{j} + \beta_{15}C2_{j}*Jrisk_{j} + u_{1j} + u_{1t} + u_{0(jt)}$$

$$\beta_{2jt}=\beta_{20} + \beta_{21}Jrisk_{j} + \beta_{22}C1_{j} + \beta_{23}C2_{j} + \beta_{24}C1_{j}*Jrisk_{j} + \beta_{25}C2_{j}*Jrisk_{j} + u_{2j} + u_{2t} + u_{0(jt)}$$

Within the context of equation (3.1) $Jrisk_{j}$ represents judge *j's* risk propensity scale score from the DOSPERT. The interpretation of this model is similar to that of equation 2.1 and 2.2. β_{10} continues to represent the average level of distinctive accuracy for the reference group while β_{11} represents the average level of distinctive accuracy at varying levels of judge risk propensity. β_{12} and β_{13} carry the same interpretation from equation 2.2. The new coefficients β_{14} and β_{15} represent the average level of distinctive accuracy given the combined influence of judge risk propensity and information relevance conditions. For normative accuracy, the coefficients β_{24} and β_{25} represent the average level of normative accuracy given the combined influence of judge risk propensity and information relevance conditions. A positive interaction would indicate that higher levels of risk-taking by the judges leads to higher accuracy. However, it is not

clear without further analysis at what value of the condition judge risk-taking is positively associated with distinctive accuracy without follow up decomposition of the interaction.

As with any analysis of multiple models within a single experimental context it is worthwhile to explore issues of type I error. Familywise error occurs when the chance of finding an effect is inflated by conducting multiple analysis using the same data and similar hypothesis. In order to control for the six models conducted with the current experiment a Bonferroni adjustment was used. The adjusted alpha level used in this experiment is .008. The Bonferroni adjustment is considered conservative, however is appropriate here because the alpha levels within the SAM estimates tend to be below .0001 (Beisanz, 2010).

Chapter IV: Results

Results

Hypothesis 1: Overall Accuracy Across Conditions

Across conditions, participants demonstrated significant levels of distinctive accuracy, b=.27, p < .0001, and normative accuracy, b=.49, p < .0001[7][8]. Also, to the *a priori* hypothesis, five separate analyses examining accuracy by DOSPERT domain (social, recreational, financial, health, ethical) were conducted. Specifically, data were set to include only items from each of the separate domains and then analyzed using equations 1.1 and 1.2. All regression coefficients for normative and distinctive accuracy were statistically significant and followed the pattern of the base model, with the exception of a negative coefficient for distinctive accuracy in the financial domain (see Table 3).

In the following results, the unstandardized regression coefficients are reported to enhance interpretation of the models. Using unstandardized coefficients allows for the interpretation of the slopes as unit changes in the predictor. For example, for the distinctive accuracy estimate of .27, the interpretation is a one unit change in the accuracy criterion score results in a .27 increase in the judge's rating for that target and item of the DOSPERT.[9]

---

[7] The lme4 package (version 1.1-10) for R (version 3.2.1) was utilized to analyze data using the SAM.
[8] Degrees of freedom for this model=243.
[9] All significant parameter estimates for each of the six models in the analysis fell below the Bonferroni corrected alpha of .008.

Table 3

Normative and Distinctive Accuracy for all Domains and by Domain

Fixed Effects	All Domains	S	R	F	H	E
Normative	.49 (.02)	.50 (.02)	.37 (.03)	.42 (.02)	.45 (.01)	.54 (.02)
Distinctive	.27 (.01)	.23 (.01)	.34 (.01)	-.10 (.03)	.32 (.02)	.04 (.01)

Note. S=Social, R=Recreational, F=Financial, H=Health, and E=Ethical. Values in the table are estimates from the model with the associated standard error. Data were analyzed using equations 1.1 and 1.2. All estimates within the model were significant at $p<.001$[10].

Hypothesis 2: Information Relevance

To test H_2 – that judges in the specific relevance condition would achieve higher distinctive accuracy than those in the broad relevance and low relevance conditions – a moderation analysis using equations 2.1 and 2.2 was used. To evaluate the two types of accuracy across groups, a series of moderation analyses were conducted using three dummy variables ($C1_=$low relevance, $C2_=$ broad relevance, $C3_=$ specific relevance) to evaluate normative and distinctive accuracy across conditions. Three dummy variables were used in two moderation analyses in order to account for all possible group comparisons.

Distinctive accuracy. Information relevance was a significant moderator of distinctive accuracy. In partial support of H_2, when compared to the low relevance condition, broad relevance information led to higher distinctive accuracy (b=.12, p < .0001). However, accuracy in the specific relevance condition did not differ significantly

[10] Socioeconomic status (SES) was calculated for judges using the Hollingshead method (Hollingshead, 1975). The inclusion of SES in all subsequent models did not change the interpretation of the results. Thus, SES was dropped from the models in favor of model simplicity.

from accuracy in the low relevance condition, b=-.03, p=.14. Additionally, judges in the specific relevance condition had significantly less distinctive accuracy than judges in the broad relevance condition, b=-.15, $p < .0001$ (see Figure 6). H_2 predicted that judges in the broad relevance condition would be more distinctively accurate than the low relevance condition, and judges in the specific relevance condition would have the highest level of distinctive accuracy. Thus, H_2 was partially supported by the data.

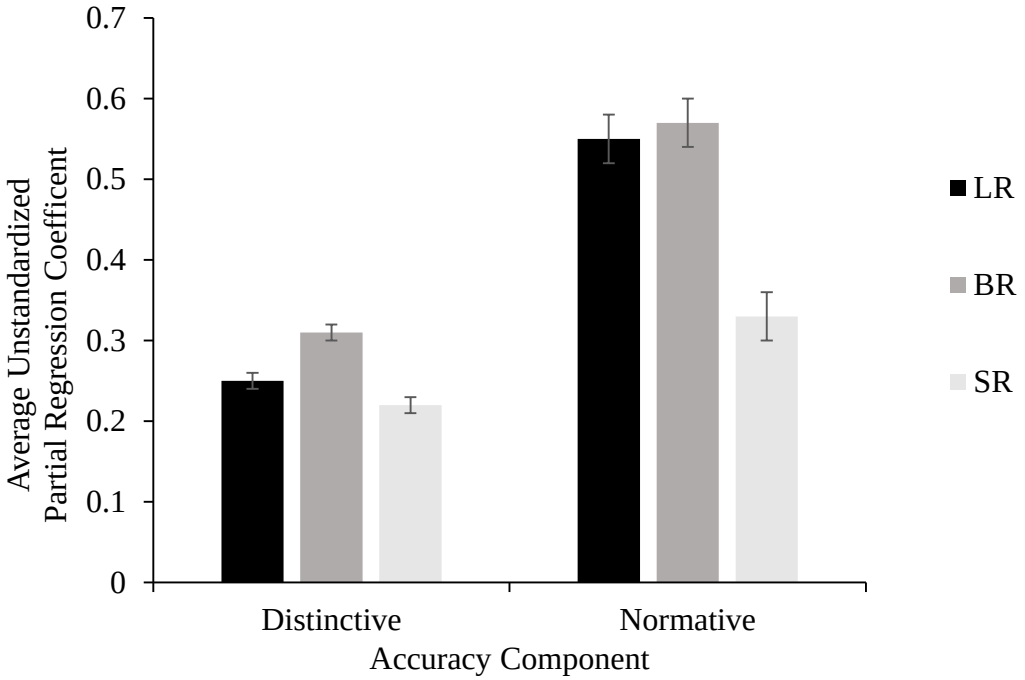

Figure 6. LR=Low Relevance, BR= broad relevance, SR=Specific Relevance. Results of analysis of the effect of information relevance on judge-target ratings, plotted by distinctive and normative accuracy for each condition. The graph shows the mean unstandardized fixed-effects partial regression coefficients, with error bars representing ± 1 *SE*.

Normative accuracy. Judges in the low relevance and broad relevance conditions did not differ in normative accuracy, b=.02, p=.41. Judges in the specific relevance

condition demonstrated significantly less normative accuracy than judges in low relevance condition, b=-.27, p < .0001. Additionally, judges in the specific relevance condition had significantly less accuracy than those in the broad relevance condition, b=-.29, p < .0001[11].

Hypothesis 3: Individual Differences in Judge Risk

To test H_3 – that judge risk propensity will moderate the relationship between information relevance condition and accuracy – a moderation analysis using equations 3.1 and 3.2 was conducted. Prior to this analysis, a simple moderation analysis of judge risk propensity on judge-target pairs was conducted. Results yielded a significant interaction between judge risk propensity and normative accuracy (b=-.003, p < .001) but not for distinctive accuracy (b=-.0009, p=.003)[12]. A negative coefficient for normative accuracy means that normative accuracy declines as judge risk scores on the DOSPERT increase. While the coefficient for normative accuracy is statistically significant, the effect size is small.

To test the second part of H_3 regarding a three-way interaction, the same dummy coding scheme from H_2 was utilized for condition, and the scale score for risk propensity from the DOSPERT was entered as a moderator for each level of the condition to create a three-way interaction. All three-way interactions for distinctive and normative accuracy were nonsignificant (ps > .32).

[11] Degrees of freedom for Hypothesis 2 models=237.
[12] Degrees of freedom for Hypothesis 3 models=234.

Three-way interactions within a multilevel model can yield results that are not statistically significant even when there is a meaningful three-way interaction. In other words, judge risk may moderate accuracy at one of the levels of information relevance, but the omnibus test could yield a non-significant statistic. Therefore, the data were subset by condition in order to conduct three separate moderator tests, with judge risk entered as a moderator of accuracy. All interactions across conditions were non-significant (*ps*>.44). This nonsignificant result was consistent with the main three-way interaction. Specifically, the influence of judge risk on accuracy was the same for each of the conditions when they were analyzed separately. Thus, the non-significance of the main three-way analysis was supported.

Chapter V: Conclusions

General Discussion

The field of the accuracy of personality judgment has established that individuals can and do make accurate judgments of personality even in zero acquaintance situations. This study sought to contribute further to the literature in three ways by 1) providing additional empirical evidence of how accurately people judge risk-taking in unacquainted individuals by extending previous methods that used pictures and unidimensional measures of risk, 2) extending previous findings on information relevance as a moderator of accuracy to include risk-taking, and 3) investigating the hypothesis that a perceiver's risk-taking propensity would moderate the relationship between information relevance and accuracy. H_1 was based on the Realistic Accuracy Model of personality judgment and specifically predicted that judges would accurately perceive trait level risk-taking regardless of condition. This hypothesis was fully supported by the data. H_2 stated that as information relevance increased from targets discussing their general personality traits, to attitudes towards risk, and finally to a concrete example of their risk-taking, that accuracy would increase. This hypothesis was partially supported by the data. While accuracy did increase from the low relevance to the broad relevance condition, the specific relevance condition led to significantly less normative and distinctive accuracy than the broad relevance condition but did not differ from the low relevance condition. H_3 stated that at different levels of information relevance, a judge's personal risk propensity would moderate accuracy. Specifically, higher risk experience would increase accuracy but more in the low relevant condition than the other conditions. H_3 was not supported by the data.

Accuracy of Risk Judgments

Across conditions, judges accurately rated target responses on the DOSPERT scale. The design used in the current experiment and accompanying findings are a theoretical replication of several experiments that have demonstrated individuals can accurately judge personality traits of others even in a situation where they have never met the person before. These findings lend further support to the utility of using the RAM to understand accuracy. The RAM is used to generate testable hypotheses and provides an organizational framework with which to understand the process of making accurate judgments. The model has been widely used in research on personality judgment, broadly speaking, and the findings here demonstrate that the use of the RAM also applies to judgments of risk.

Previous research into the accuracy of judging risky behaviors (e.g., violent behavior, criminal activity) has demonstrated these behaviors are accurately percieved from pictures of faces. The experiment presented here uses 1-minute video clips as the stimulus materials, rather than a picture format. Additionally, previous research paradigms have used a unidimensional measurement of risk-taking (e.g., "How risky is this person?"). Findings from the current experiment show that using video stimuli and a multidimensional measure of risk produce significant levels of accuracy. The use of videos and a broader multidimensional measure of risk are notable from a methodological perspective as an extension. Using video stimuli more closely resembles how people make judgements about personality on a daily basis. Video clips have dynamic body language, facial expression, tone of voice, and other cues that pictures do not possess. Further, the use of the DOSPERT gives information not only about multiple domains of

risk-taking, but also yields information about risk propensity and attitudes across those domains. Asking individuals to rate a person's riskiness using a single questions (e.g., "How risky is this person?") makes sense given participants must rate 200 plus pictures in an experimental session, however this single-question approach does not adequately capture the complexity of risk-taking.

As an additional methodological point, previous experiments have exclusively studied extreme risk behaviors like violence or criminal behavior. The current study used the DOSPERT as a measure of risk propensity and risk attitudes. The items of the DOSPERT represent not only a dimensional account of risk but also describe behaviors everyday people are likely to engage in. Therefore, the findings provide entirely new data demonstrating that accuracy occurs for common and frequent risk behaviors. Additionally, it is also of note that judges did not know the participants. With no prior acquaintanceship with the target and only 60 seconds of video exposure, judges were able to perceive the propensity and attitudes of the seven targets with a statistically significant level of accuracy.

Last, accuracy is defined by approximation to an actual or real value. The RAM posits that personality traits are real and thus have a distinct value, however this value cannot be known except through the use of multiple criteria. A person's self-report of their traits is one channel of information that can be used to gain closer approximation to the person's actual trait level. Adding additional sources of information including acquaintance ratings of the person's personality, or behavioral outcomes related to the trait, are more channels that increase an approximation to the true value of the trait. Accuracy is also best achieved by examining real people, in real situations. However,

previous studies are based on perceptions of risk instead of accuracy. These investigations into accuracy have simply asked for a judge's perception of the riskiness of the target by asking how risky is the person but have not asked about how that person is across real day-to-day situations (i.e., DOSPERT items). The current experiment uses an established method for creating an accuracy criterion along with the SAM to model both normative and distinctive accuracy. Overall, these findings provide an improved understanding of the accuracy of risk judgments built on established and empirically validated models of accuracy.

Information Relevance

While several studies have previously shown the impact of information quality and quantity on accuracy, this moderator from the RAM remains understudied. The question remains what types of information are the highest quality and thus most diagnostic of particular traits. Previous research has demonstrated that thoughts and feelings provided by the target yield differential effects on accuracy compared to behaviors (Andersen & Ross, 1984; Beer & Brooks, 2011; Letzring & Human, 2013). The current research has begun to explore this topic, but no previous research has attempted to elucidate what types of information are of high quality when judging risk-taking. Hypothesis 2 specifically deals with addressing these questions by varying the relevance of the information provided to judges. Therefore, the second hypothesis was that accuracy would be greater as the relevance or specificity of information increased from each condition. Specifically, the hypothesis was that judges who watched the seven targets discussing a particular example of their risk-taking behavior would more accurately judge those targets than judges in the broad relevance and low relevance

conditions. The data partially supported the hypothesis as accuracy did increase from low relevance to broad relevance conditions, but the high relevance condition had less normative and distinctive accuracy than the other two conditions.

Judges in the specific relevance condition judged targets as accurately as judges in the low relevance condition. This finding suggests that the influence of watching someone tell a specific example of risk-taking from his or her life is as diagnostic for cue utilization as someone talking broadly about his or her personality. The question is why would highly relevant information yield approximately equal accuracy to discussing general personality traits? A plausible explanation for this finding is judges focused on the specific risk story and overgeneralized the targets risk-taking across domains. Focusing on the story may lead to an accuracy within the domain itself, but less accuracy in the other domains as the judge attempts to use the information from the risky story to estimate risk-taking in other areas.

Consider the target who described taking extreme risks riding his motorcycle but who scored low on all other risk domains. Further, targets were asked to give an example of the riskiest thing they had ever done, and all of the examples occurred several years ago. A simple explanation for the effect of condition on distinctive accuracy is that these stories are not currently diagnostic of the targets risk-taking. In fact, a significant amount of research on risk-taking has shown that risk propensity decreases with age and maturity (Byrnes, Miller, & Schafer, 1999; Deakin, Aitken, Robbins, & Sahakian, 2004; Figner, Mackinlay, Wilkening, & Weber, 2009). In other words, targets told stories of youthful risk-taking but have reduced their risk propensity and adjusted their risk attitudes since.

Therefore, judges in the specific relevance condition would perceive target risk-taking with significantly less accuracy than the broad relevance or low relevance conditions.

It is interesting that the effect of listening to a story about a risky behavior resulted in the same pattern in both normative and distinctive accuracy, although significance across information relevance did differ between the two. A target discussing previous risk-taking activities should not affect a judge's normative accuracy through the aforementioned past-present incongruence, as that effect is specific to distinctive judgments of the target. However, it is plausible that the information available in the specific relevance condition elicited an overestimation of risk-taking as was suggested for distinctive accuracy. Specifically, watching descriptions of the riskiest behavior someone has engaged in provides an anchor point for estimating the average risk across items and domains. If the anchor point is greater than the normative profile, then accuracy in the specific relevance condition would be lowest among the three conditions.

In fact, in the judgment and decision-making literature, several studies have shown that estimating the outcome or riskiness of a decision can be tied to an anchor or set point (Epley & Gilovich, 2001; Epley & Gilovich, 2006; Northcraft & Neale, 1987). This research shows that individuals make estimates of probabilities by first setting an anchor point and then adjusting to and from that point. The person making the judgment can set the anchor point but often the anchor comes from the environment (i.e., used car salespersons make an initial offer, and the buyer negotiates around that price). Anchoring-adjusting is applicable here as a specific story about the riskiest behavior a person has engaged in sets the anchor at above the norm by definition and like estimating outcomes judges in the current experiment must estimate risk level. One counter-

argument is that the targets in the current experiment did not disclose particularly risky activities but instead stuck to relatively typical risks. The previous argument is not valid however as each of the targets stories represent risk propensity and attitudes that score high in the DOSPERT and thus are above the normative scores.

The primary concern of the current experiment is the influence of information relevance on distinctive accuracy. The findings were not consistent with previous research that found distinctive accuracy was similar between information about thoughts and feelings and information about behaviors. Letzring and Human (2013) found that individuals judged on personality traits after discussing their thoughts and feelings in comparison to discussing their behaviors had similar levels of distinctive accuracy, but both conditions had higher distinctive accuracy compared to overserving the targets' behavior (i.e., playing Jenga). One possible explanation for this difference lies in the type or content of the information across conditions. In the current experiment targets were asked in an open-ended format to discuss their attitudes towards risk, while Letzring and Human asked participants to discuss their thoughts and feelings in different situations (i.e., with family or at work). The difference is that attitudes towards risk are highly specific to the trait (risk) that participants were asked to judge. Targets each discussed different attitudes towards risk across domains, which yielded information directly relevant to the trait for making accurate distinctive perceptions. As mentioned previously, targets in the specific relevance condition provided cues directly related to a domain within risk, which could have led to over estimating across risk domains. Given this possibility, it is also important to note that judges in previous studies have had significantly more time (3, 5, or 30 min.) to acquire information. It is possible that the

disadvantage demonstrated in these findings for watching targets discuss their behaviors disappears with a greater quantity of information because they may be exposed to cues relevant to different risk domains given more time.

While the current experiment was primarily interested in distinctive accuracy, it is informative to consider normative accuracy as well. Judges were more normatively than distinctively accurate, which replicates previous findings for traits (Biesanz, 2010). This means that judges were more accurate at judging the general risk characteristics across domains than they were at judging distinctive differences in the targets from the normative profile. This finding is not surprising considering the amount of empirical evidence that demonstrates people are accurate at judging the average person (e.g., Biesanz, 2010; Human & Biesanz, 2013; Letzring & Human, 2013). Additionally, the pattern of results for normative accuracy in the current experiment are similar to distinctive accuracy in that both low relevance and broad relevance conditions yielded more accuracy than the specific relevance condition. However, the low relevance and broad relevance conditions did not produce significantly different levels of normative accuracy. Judges did not appear to gather more quality information in the broad relevance condition and thus information about general personality compared to information about their broad relevance attitudes yielded the same benefit towards normative accuracy.

Judge Risk Moderation of Information Relevance

The third hypothesis specifically deals with the interaction between the good judge and information quality moderators. According to the RAM (Funder, 1995), this interaction is referred to as *sensitivity* and describes how a judge's personal traits or individual differences interact with the availability of relevant information. Broadly

speaking, when judges are high on a trait, they are sensitive to detecting information relevant to that trait. Thus, the third hypothesis is that a judge's risk propensity score on the DOSPERT would moderate accuracy and specifically as information relevance changed. This two-part hypothesis first stated that judge risk would moderate distinctive accuracy regardless of the condition, while the second part stated that when information relevance is low, judges with high risk propensity would demonstrate a higher degree of distinctive accuracy when compared to judges with lower risk propensity. Further, this advantage would dissipate as the availability of higher quality information increased across condition. The data did not support this hypothesis. There was not a significant interaction between judge risk and distinctive accuracy. While there was a statistically significant interaction of judge risk propensity and normative accuracy, the size of the beta weight is so small as to render a prediction with functionally zero predictive power. The three-way interaction between judge risk, information relevance, and judge-target ratings were all nonsignificant.

The exact explanation for why judge risk did not interact with information relevance and accuracy is elusive. On the surface, this conclusion seems to contradict the RAM prediction that increased judge sensitivity should lead to increased accuracy, however upon deeper reflection it is evident that these findings are consistent with the RAM. First, the RAM explicitly points out that judge sensitivity is specific rather than general and, as a result, judges are more sensitive and place more weight on specific or specific relevance information. In this context, it is not clear, nor can it be from the design of the current experiment, whether higher self-report scores of risk propensity necessarily mean greater sensitivity to risk-taking in others. Risk propensity as measured

by the DOSPERT is the likelihood a person would engage in a given activity across items and domains. Sensitivity is related to the detection and utilization stage of accuracy judgments; it is possible that being more likely to participate in risky behaviors affords an advantage in detecting cues from others but not necessarily effectively use those cues. Further, simply being more willing to engage in risky behaviors does not mean the judge has a significant amount of personal experience with risk-taking.

Another possible explanation for the null findings in the current experiment relates to the potential for a hidden moderator (e.g. judge social intelligence) to the relationship between judge risk and accuracy. The current experiment was the first to examine judge risk propensity as a moderator of accuracy and thus little is known about the relationship of this potential moderator. One explanation for the findings here is that judge risk propensity simply does not moderate information relevance and accuracy, in contrast another explanation is that this relationship may be complex. In other words, it is possible that another variable not accounted for in the current experiment is influencing the relationship. Take cue detection for instance: because cue detection involves mental processes, individual differences in mental ability can affect accuracy. Assuming judge risk propensity is related to accuracy, it is possible that not all people are equal in their capacity to utilize cues for judgments.

For example, Christensen et al. (2005) demonstrated that mental ability was the best predictor of accuracy in a video paradigm similar to the one used in the current experiment. Thus, mental ability may moderate the relationship between judge risk propensity and accuracy of risk judgments. Mental ability is not the only possible hidden moderator however, as nonverbal sensitivity has also been shown to affect accuracy

(Ambady, LaPlante, & Johnson, 2001; Funder & Harris, 1986). Individuals show a global ability to detect nonverbal cues that varies from person to person. It is reasonable to argue that the sample from the current experiment has a consistent normal distribution of this ability, but consistency between judge risk propensity scores and nonverbal sensitivity in each judge is not implied. Therefore, this possible incongruence of individual differences may lead to a moderator to judge risk propensity and accuracy with or without information relevance.

Implications

Theoretical. First and foremost, the current experiment was a test of the broad question of the degree to which people can accurately judge trait risk-taking in others in a zero acquaintance circumstance. In this endeavor, the results give direct empirical evidence that people achieve both normative and distinctive accuracy of trait risk-taking in others. Further, predictions using the RAM yielded some results consistent with the moderators of accuracy and spelled out by the RAM. The finding has important implications since, before this study, no experiment had utilized the RAM or the accompanying statistical method SAM to investigate the accuracy of risk judgments in others. The results establish that both normative and distinctive accuracy for risk propensity and risk attitudes is possible and interpretable according to the RAM and SAM. Since the methods of the current experiment used targets that judges did not know, and each judge only watched a 60-second video clip, it is clear from the findings that accuracy occurs within the context of limited information and no background knowledge.

Previous research on risk judgments of others has primarily used pictures and written vignettes. The current experiment introduces a new method for stimulus materials

by utilizing a well-established paradigm of using video clips of targets. This addition is significant because while the judgment of pictures yields theoretically useful results, the use of videos improves on that process in two important ways. First, videos are more ecologically valid than still pictures because much of our personality judgments occur in social situations where individuals speak and move. Second, and related to the first, the ecological nature of videos is relevant since a wealth of additional cues and information, as well as potential misinformation, becomes available in these social situations.

Additionally, only one previous study has attempted to assess the accuracy of judging risk in others (Mishra & Sritharan, 2012). They used a single question asking judges how risky they thought each person in the photograph was. The current experiment expanded on that single item measure using the SAM to model accuracy not only across several domains of risk (e.g., recreational or financial) but provided additional information about normative and distinctive accuracy. This latter aspect is especially important as Biesanz (2110) has pointed out that with the contributions of normative and distinctive components, accuracy of person judgments is better understood. Specifically, most people are average on where they fall on any given trait, and thus, measurement of accuracy must account for that approximation, but we are also interested in how well people can judge how a single individual is above, below, or the same as the normative profile. The SAM accounts for these differences, and the findings are consistent with predictions from the SAM. The results also indicate that, as with other findings from the field, normative accuracy tends to be higher than distinctive accuracy. The findings add to the literature and supports current interpretations using the RAM and SAM.

Risk is a complex phenomenon and is conceptualized using a dimensional perspective. Previous research into risk has typically utilized a unidimensional approach to looking at single behaviors such as criminal behavior or gambling. Also, these paradigms tend to focus on behavior considered to be atypical of the average person (i.e., criminal behavior). The design used here expands the knowledge base by using a multidimensional measurement of risk propensity and risk attitudes, thus providing a measure that is more free of inflation effects caused by grouping all risk-taking into one unidimensional construct. Accuracy was achieved across measurements of target risk propensity and attitudes. This finding implies accuracy is not an artifact of simple guessing at the risk propensity based on individual cues and stereotypes. The use of more germane and valid ecological behaviors in the DOSPERT (e.g., disagreeing with friends, challenging authority) also makes the findings from this study more generalizable to the population. At the same time, the DOSPERT includes unethical behaviors and behaviors looked at negatively. Taking together, the more acceptable risk behaviors with the negative responses leads to a more accurate picture of risk the average person undertakes. Therefore, the judges in the current experiment were carrying out judgments individuals would encounter in day-to-day life.

Applied. The findings of the current study have potential application across several settings, including business, security, and personal life. Beyond the finding that individuals can accurately judge risk, the results from the current experiment can inform what kinds of information may be most useful in increasing both normative and distinctive accuracy. A focus on thoughts, feelings, or attitudes seems to yield the highest accuracy, with specific risky behaviors yielding the least accuracy. Given that accuracy

of risk judgment is possible but modified by the quality of information, managers or human relations staff could be trained with these findings in mind. A hiring manager could focus precious initial interview time on attitudes versus examples of risk-taking. In comparison, security personnel charged with determining the risk level of people with whom they interact could benefit from an emphasis on attitudes and general personality in investigations, especially when they have limited information or limited contact with the target.

The efficent use of cues and accuracy is especially important in jobs where risk-taking and assessment are a significant aspect of the role, as in managing growth stock portfolios in finance. In these cases, individuals utilize risk assessment and estimation daily and sometimes in very short amounts of time. It is, therefore, critical that employees be screened and hired with their risk-taking in mind. The issue at hand is not whether an employee will risk the company's money but that they may engage in risk behaviors that introduce risk complexity into their behavior (i.e. driving drunk, gambling). Based on null results from the current experiment, employees with higher risk propensity within the context of the job may not have an advantage over their lower risk propensity peers. The implication here is not to replace already well-established methods of screening individuals for risk work, but to enhance and augment current practices to identify the risk-taking profile of potential employees and improve human resource management. Further, these findings support the practice of using a semi-structured interview in which employers ask employees to describe their thoughts and attitudes towards certain job-related aspects.

Based on the null findings from hypothesis 3, it is interesting to note that a judge's personal risk propensity may not be a good predictor of how well they will accurately judge risk-taking in others. This implication is made with interpretive caution, however, as further studies should be conducted to investigate the limitations with these findings discussed herein.

In personal life, individuals are constantly interacting with new people by making new friends, going on dates, meeting new co-workers or bosses or just interacting with individuals on the street. The implications from the current study are that when judging the risk propensity and attitudes of a potential new friend or co-worker, individuals may find it more beneficial to focus on gathering information about thoughts towards risk and take behavioral examples of risk with a grain of salt.

Limitations

First, the current findings are based on data from a convenience sample of college students in Southeastern Idaho taking an introductory psychology class. Several potential limitations to the external validity of the findings occur from this particular sample. First, college samples are mostly made up of younger individuals. There is a possibility that age or life experience could significantly increase accuracy, but with an age-restricted sample it is not possible to examine this moderator. Also, 18- to 25-year-old college students in rural Idaho may not generalize well to the broader population. The broader population tends to be older and have more life experience, and it may be the case that older individuals are more adept at detecting or utilizing cues for making risk judgements in others. College populations themselves engage in riskier social behaviors, which may cause them to overestimate the riskiness of others.

A further limitation of this study involves the creation of stimulus materials using videos instead of live interactions. While live interactions are more ecologically valid than videos, the procedure employed in the current experiment has significant importance. The main reason live interactions may be preferred is the exchange of information and processing of cues are a dynamic process between a judge and target. A person meets someone for the first time and during a short conversation tries to assess their personality. During that conversation, both individuals are involved in processing information and monitoring and adjusting their own responses and behaviors to one another. This is fundamentally different than watching a target in a quiet laboratory setting on a computer.

Despite the preference for live interactions, the use of video stimuli are valid and established methods for the following reasons. First, using videos allows for control of variables that occur in live interactions. To understand information quality, it is important to control extraneous variables. Using videos also allows judges to watch several targets in one session, which aids in isolating the effect of information quality as studied here and increases the reliability of estimations for accuracy scores. Next, while videos are not as ecologically valid as a live interaction, they are an improvement of ecological validity over still pictures used in all previous research on the accuracy of risk judgments. Last, it could be argued that 60-second videos are also not ecologically valid and could affect the results of the findings. However, it was the purpose of this study to investigate accuracy under circumstances of very limited information and as such this apparent limitation is purposeful. It would be interesting to replicate the current experiment with a live interaction to capture the dynamic nature of a “getting to know you” scenario, which is

arguably more ecological. Further, it would be interesting to manipulate the time of the videos by using different video lengths and see what effect information quantity plays in accuracy and the accompanying moderators.

Future Directions

Research going forward should continue to establish the cues and mechanisms underlying accuracy of risk judgment. For example, it is interesting to consider how normative and distinctive accuracy would be affected if dimensions the judges rate on the risk measurement are matched with the target's story (e.g., recreational risk-taking) and then to compare that accuracy to other dimensions. Perhaps distinctive accuracy overall would suffer, but would increase in the dimension-congruent circumstances. Further, the cues themselves are not well understood. Within a single 60-second video there is a wealth of verbal, nonverbal, and stereotype (clothes or hairstyles) information available to a potential judge. What is not currently understood is the relative weights of the cues and their relationship to other cues in judging accuracy. Finally, the mechanisms of accuracy need further elucidation. The current experiment shows people do accurately judge trait risk, and that information quality moderates that relationship. Consider a situation where making a judgment comes under a time constraint or divided attention. In these cases, working memory ability would significantly impact accuracy.

In the case of deception, accuracy could be affected by intentional deception on the targets part but may be moderated by any of the variables from the RAM model. Many cues available for detection and used for judgements are nonverbal. Deceiving others requires the control and management of nonverbal behaviors while maintaining the deception (i.e., not talking too fast or making eye contact). It is not currently known how

the process of deceiving could interfere or confuse commonly used nonverbal cues for perceiving personality. A possible practical application is in circumstances where deception is practiced regularly to cover up risk attitudes or behaviors (e.g., criminal interrogation). In these circumstances, accuracy may be improved by isolating the effects of willful deception. Future research should investigate accuracy of personality and risk judgements under scenarios of deception. Specifically, targets could be induced to cover up or deceive a peer or interviewer about an actual trait or attitude. Research into the impact of deception on accuracy could yield important information about daily relationships and further enlighten theoretical investigations of moderators of accuracy.

Another possible path forward is to examine the accuracy of risk judgments in populations that specifically deal with risk (e.g., first responders, security personnel, military, or nuclear facility workers). People who deal with risk have developed a higher degree of accessibility to risk concepts and schemas compared to individuals in low or no risk careers. The chronic accessibility that comes from constantly assessing risk in the environment and in other people may lead to higher sensitivity to risk cues and increased distinctive accuracy. It would also be interesting to examine the effect of acquaintanceship in this population. Individuals who work together for more than six months in a high-risk job might be able to more accurately estimate risk in their coworkers. The question is whether possible moderators like experience on the job, leadership experience, or age play a role in accuracy. On this note, researchers would also benefit from investigating how accuracy in these high-risk careers interacts with organization behavior. Specifically, how do workers utilize information about risk in others within an organizational framework?

Finally, in the process of studying accuracy, it is important to acknowledge that accuracy itself is an ideal and to know the actual value of someone on any given trait comes only through approximation. As multiple channels of information are added together and aggregated, one is closer to reality than with no information. To set an accuracy criterion, simple self-report is considered fraught with problems. It is a significant improvement and thus closer approximation to reality to include information from acquaintances along with a target's self-report. One way to accomplish the task of increased approximation is to examine specific domains of risk-taking and collect behavioral data from those domains to improve the accuracy criterion. This could be accomplished by using experience sampling or archival data about risk behaviors (e.g., job related accidents, reprimands for risk-taking at work). In a laboratory setting, the social domain of risk-taking could be examined by having participants engage in a significant interaction with peers and code their behavior of social risk-taking (e.g., sharing private feelings, disagreeing with the group). Adding a layer of information about actual behavioral outcomes - while complex, costly, and time-consuming - is a worthwhile endeavor. Future research on this topic should labor to include behavioral measures of risk through experience sampling techniques or objective measures of risk-taking (worker safety record).

Conclusion

The results of this study provide important empirical evidence of the accuracy of judging risk in others. First, previous research on accuracy had focused on the Big Five personality traits and research that did investigate trait risk focused on single-dimension measures and the use of still photographs. This study has contributed to the literature by

demonstrating that people can accurately judge both the normative and distinctive aspects of trait risk characteristics while using empirically sound models and methods to understand these judgments. Second, the current experiment explored the impact of information quality on the accuracy of risk judgment in others. The current experiment demonstrated that listening to broad attitudes towards risk yields the highest distinctive accuracy, while listening to specific risk behaviors yields the lowest accuracy. The novel contribution comes in providing evidence for an effect of information quality in risk judgments versus personality judgments and the use of information that directly matches the perceived trait. Finally, a judge's personal trait risk did not moderate accuracy or the interaction between information quality and accuracy. This finding is important because it highlights the need to provide further empirical evidence of accuracy paradigms and suggests that trait elements beyond mere risk propensity may be crucial to improving accuracy.

References

Allais, M. (1979). The so-called Allais paradox and rational decisions under uncertainty. In *Expected utility hypotheses and the Allais paradox* (pp. 437-681). Springer Netherlands.

Allman, W. F. (1985). Staying alive in the 20th century. *Science, 85*, 30-37.

Allport, F. H., & Allport, G. W. (1921). Personality traits: Their classification and measurement. *Journal of Abnormal and Social Psychology, 16*, 6-40.

Andersen, S. M., & Ross, L. (1984). Self-knowledge and social inference: I. The impact of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology*, *46*, 280. http://dx.doi.org/10.1037/0022-3514.46.2.280

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, *69*, 518-529. doi:10.1037/0022-3514.69.3.518

Appelt, K. C., Milch, K. F., Handgraaf, M. J., & Weber, E. U. (2011). The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making, 6*, 252-262.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9-30. doi:10.1111/1468-2389.00160

Beer, A., & Brooks, C. (2011). Information quality in personality judgment: The value of personal disclosure. *Journal of Research in Personality, 45*, 175-185. doi:10.1016/j.jrp.2011.01.001

Bell, D. E. (1995). Risk, return, and utility. *Management Science, 41*, 23-30.

Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research, 45*, 853-885. doi:10.1080/00273171.2010.519262

Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science*, *21*, 589-594. doi:10.1177/0956797610364121

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, 92, 119. http://dx.doi.org/10.1037/0022-3514.92.1.119

Blais, A. R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1*, 33-47.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology, 34*, 164-181. doi:10.1006/jesp.1997.1347

Bontempo, R. N., Bottom, W. P., & Weber, E. U. (1997). Cross-cultural differences in risk perception: A model-based approach. *Risk Analysis, 17*, 479-488. doi:10.1111/j.1539-6924.1997.tb00888.x

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645. http://dx.doi.org/10.1037/0022-3514.62.4.645

Brunswik, E. (1952). *The Conceptual Framework of Psychology.* Chicago, IL: University of Chicago Press.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk-taking: A meta-analysis. *Psychological bulletin*, *125*(3), 367. http://dx.doi.org/10.1037/0033-2909.125.3.367

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20,* 1194–1198. doi:10.1111/j.1467-9280.2009.02423.x

Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance*, *18*, 123-149. doi:10.1207/s15327043hup1802_2

Cline, V. B., & Richards Jr, J. M. (1960). Accuracy of interpersonal perception: A general trait?. *The Journal of Abnormal and Social Psychology*, 60, 1. http://dx.doi.org/10.1037/h0041320

Cronbach, L. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52*, 177. doi:10.1037/h0044919

Crow, W. J., & Hammond, K. R. (1957). The generality of accuracy and response sets in interpersonal perception. *The Journal of Abnormal and Social Psychology, 54*(3), 384. http://dx.doi.org/10.1037/h0042990

Colvin, C. R. (1993). Judgable people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, *64,* 861-873.

Cumming, G. (2013). The new statistics why and how. *Psychological Science*, *25,* 7-29. doi:10.1177/0956797613504966

Deakin, J., Aitken, M., Robbins, T., & Sahakian, B. J. (2004). Risk-taking during decision-making in normal volunteers changes with age. *Journal of the International Neuropsychological Society*, *10*(04), 590-598. http://dx.doi.org/10.1017/S1355617704104104

Einhorn, H. J. (1970). The use of nonlinear noncompensatory models in decision making. *Psychological Bulletin, 73,* 221–230. doi:10.1037/h0028695

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12,* 391-396. doi: 10.1111/1467-9280.00372

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic Why the adjustments are insufficient. *Psychological science, 17*, 311-318. doi: 10.1111/j.1467-9280.2006.01704.x

Erdheim, J., Wang, M., & Zickar, M. J. (2006). Linking the Big Five personality constructs to organizational commitment. *Personality and Individual Differences, 41*, 959-970. doi:10.1016/j.paid.2006.04.005

Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology*, *33*, 217-236. doi:10.1037/h0058565

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: age differences in risk-taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 709. http://dx.doi.org/10.1037/a0014983

Fishburn, P. C. (1982). The foundations of expected utility. *Theory & Decision Library*. 31, 176.

Fleeson, W., & Noftle, E. E. (2009). In favor of the synthetic resolution to the person–situation debate. *Journal of Research in Personality*, 43, 150-154. doi:10.1016/j.jrp.2009.02.008

Flink, C., & Park, B. (1991). Increasing consensus in trait judgments through outcome dependency. *Journal of Experimental Social Psychology*, *27*, 453-467. doi:10.1016/0022-1031(91)90003-O

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102,* 652-670. doi:10.1037/0033-295X.102.4.652

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*, 177-182. doi:10.1177/0963721412445309

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, *55*, 149-158. doi:10.1037/0022

Funder, D. C., & Harris, M. J. (1986). On the several facets of personality assessment: The case of social acuity. *Journal of Personality*, *54*, 528-550. doi: 10.1111/j.1467-6494.1986.tb00411.x

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, 14, 483-488. doi: 10.1177/0963721412445309

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality, 42*, 1476-1489. doi:10.1016/j.jrp.2008.06.013

Haselton, M. G., & Funder, D. (2006). *The evolution of accuracy and bias in social judgment*. In M. Schaller, D. T. Kenrick, & J. A. Simpson (Eds.), *Evolution and social psychology* (pp. 15−37). New York: Psychology Press.

Hastie, R. (2001). Problems for judgment and decision making. *Annual review of psychology*, *52*, 653-683. doi: 10.1146/annurev.psych.52.1.653

Highhouse, S. (2001). Judgment and decision-making research: Relevance to industrial and organizational psychology. In N. Anderson, D. S. Ones, H. P. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 2, pp. 314–331). London: Sage.

Howell, D. C. (2010) *Statistical methods for psychology* (7th ed.). Belmont, CA: Cengage Wadsworth.

Hollingshead, A. A. (1975). *Four-factor index of social status*. Unpublished manuscript, Yale University, New Haven, CT.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences*, *32*, 401-414. doi:10.1016/S0191-8869(01)00032-0

Human, L. J., & Biesanz, J. C. (2013). Targeting the good target an integrative review of the characteristics and consequences of being accurately perceived. *Personality and Social Psychology Review*, *17*, 248-272. doi:10.1177/1088868313495593

Hunt, R., Krzystofiak, F., Meindl, J., & Yousry, A. (1989). Cognitive style and decision making. *Organizational Behavior and Human Decision Processes, 44*, 436–453. doi:10.1016/0749-5978(89)90018-6

Isen, A. M., Nygren, T. E., & Ashby, F. G. (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. *Journal of*

Personality and Social Psychology, *55*(5), 710-717. doi:10.1037/0022-3514.55.5.710

Isen, A. M., & Patrick, R. (1983). The effect of positive feelings on risk-taking: When the chips are down. *Organizational behavior and human performance*, 31, 194-202. doi:10.1016/0030-5073(83)90120-4

Jarvis, B. G. (2008). MediaLab (Version 2008.1.33) [Computer Software]. New York, NY: Empirisoft Corporation.

Jia, J., Dyer, J. S., & Butler, J. C. (1999). Measures of perceived risk. *Management Science*, *45*, 519-532. doi:10.1287/mnsc.45.4.519

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.). *Handbook of personality: Theory and research*. 3rd ed. (pp. 114-158). New York: Guilford.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61, 521-551. doi:10.1111/j.1467-6494.1993.tb00781.x

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*, 765-780. doi:10.1037/0021-9010.87.4.765

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology, 87,* 530-541. doi:10.1037/0021-9010.87.3.530

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives, 5,* 193-206.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society, 47,* 263-291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American psychologist, 39,* 341-350. doi:10.1037/0003-066X.39.4.341

Kirby, K. N., & Maraković, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic bulletin & review, 3,* 100-104. doi:10.1111/j.1467-6494.1996.tb00513.x

Kogan, N., & Wallach, M. A. (1964). *Risk-taking: A study in cognition and personality*. New York: Holt.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64,* 311-337. doi:10.1111/j.1467-6494.1996.tb00513.x

Leikas, S., Lönnqvist, J. E., & Verkasalo, M. (2012). Persons, situations, and behaviors: Consistency and variability of different behaviors in four interpersonal

situations. *Journal of Personality and Social Psychology*, *103*, 1007-1022. doi:10.1037/a0030385

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality, 42*, 914-932. doi:10.1016/j.jrp.2007.12.003

Letzring, T. D. (2010). The effects of judge-target gender and ethnicity similarity on the accuracy of personality judgments. *Social Psychology*, *41*, 42-51. doi:10.1027/1864-9335/a000007

Letzring, T. D. (in press). Observer judgmental accuracy of personality: Benefits related to being a good (normative) judge. *Journal of Research in Personality*. doi:10.1016/j.jrp.2014.05.001

Letzring, T. D., & Human, L. J. (2013). An examination of information quality as a moderator of accurate personality judgment. *Journal of Personality, 82*, 440-451. doi:10.1111/jopy.12075

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, *91*, 111-123. doi:10.1037/0022-3514.91.1.111

Levin, I. P., & Hart, S. S. (2003). Risk preferences in young children: Early evidence of individual differences in reaction to potential gains and losses. *Journal of Behavioral Decision Making*, *16*, 397-413. doi:10.1002/bdm.453

Libby, R., & Fishburn, P. C. (1977). Behavioral models of risk-taking in business decision: A survey and evaluation. *Journal of Accounting Research, 15*, 272-292.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*, 267-286. doi:10.1037/0033-2909.127.2.267

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood physical attractiveness and accuracy in first impressions of personality. *Psychological Science*, 21, 1777-1782. doi: 10.1177/0956797610388048

MacCrimmon, K. R., & Wehrung, D. A. (1986). *Taking risks: The management of uncertainty*. New York, NY: Free Press.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Routledge.

Markowitz, H. M. (1959). *Portfolio selection*. New York, NY: Wiley

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86-92. doi:10.1027/1614-2241.1.3.86

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709-734. doi:10.5465/AMR.1995.9508080335

Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology, 55*, 1–22. doi:10.1146/annurev.psych.55.042902.130709

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*, 246-268. doi:10.1037/0033-295X.102.2.246

Mishra, S., Lalumière, M. L., & Williams, R. J. (2010). Gambling as a form of risk-taking: Individual differences in personality, risk-accepting attitudes, and behavioral preferences for risk. *Personality and Individual Differences, 49*, 616-621. doi:10.1016/j.paid.2010.05.032

Mishra, S., & Sritharan, R. (2012). Personality and behavioral outcomes associated with risk-taking are accurately inferred from faces. *Journal of Research in Personality, 46*, 760-764. doi:10.1016/j.jrp.2012.06.007

Mongin, P. (1997). Expected utility theory. *Handbook of Economic Methodology*, 342-350.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin, 35*, 1661-1671. doi:10.1177/0146167209346309

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational behavior and human decision processes, 39*, 84-97. doi:10.1016/0749-5978(87)90046-X

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior,* 34, 83–11. doi:10.1007/s10919-009-0082-1

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, 55, 1−81. doi:10.1111/j.1744-6570.2002.tb00103.x

Powell, D. M., & Goffin, R. D. (2009). Assessing personality in the employment interview: The impact of training on rater accuracy. *Human Performance, 22*, 450-465. doi:10.1080/08959280903248450

Raub, W., & Snijders, C. (1997). Gains, losses, and cooperation in social dilemmas and collective action: The effects of risk preferences. *Journal of Mathematical Sociology, 22*, 263-302. doi:10.1080/0022250X.1997.9990204

Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org.

Sarin, R. K., & Weber, M. (1993). Risk-value models. *European Journal of Operations Research, 70,* 135-149. doi:10.1016/0377-2217(93)90033-J

Schoemaker, P. J. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature, 20,* 529-563.

Schoemaker, P. J. H. (1990). Are risk-attitudes related across payoff domains and response modes? *Management Science, 36,* 1451-1463. doi:10.1287/mnsc.36.12.1451

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of Management Review, 17,* 9-38. doi:10.5465/AMR.1992.4279564

Sitkin, S. B., & Weingart, L. R. (1995). Determinants of risky decision-making behavior: A test of the mediating role of risk perceptions and propensity. *Academy of Management Journal, 38,* 1573-1592. doi:10.2307/256844

Slovic, P. (1998). Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. In M. H. Bazerman, D. M. Messck, A. E. Tenbrunsel, & K. A. Wade-Benzoni (Eds.) *Environment, ethics and behavior: The psychology of environmental valuation and degradation*, pp. 277-313. San Francisco: New Lexington Press.

Slovic, P. E. (2000). *The perception of risk*. London, England: Earthscan Publications.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk analysis, 24,* 311-322. doi:10.1111/j.0272-4332.2004.00433.x

Stewart Jr, W. H., & Roth, P. L. (2001). Risk propensity differences between entrepreneurs and managers: A meta-analytic review. *Journal of Applied Psychology, 86,* 145-153. doi:10.1037/0021-9010.86.1.145

Stillman, T. F., Maner, J. K., & Baumeister, R. F. (2010). Thin slice of violence: Distinguishing violent from nonviolent sex offenders at a glance. *Evolution and Human Behavior,* 31, 298–303. doi:10.1016/j.evolhumbehav.2009.12.001

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297-323. doi:10.1007/BF00122574

Valla, J. M., Ceci, S. J., & Williams, W. M. (2011). The accuracy of inferences about criminality based on facial appearance. *Journal of Social, Evolutionary, and Cultural Psychology*, 5, 66–91. doi:10.1037/h0099274

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98, 281-300. doi:10.1037/a0017908

Vlek, C, & Stallen, P. J. (1980). Rational and personal aspects of risk. *Acta Psychologica, 45*, 273-300. doi:10.1016/0001-6918(80)90038-4

Weber, E. U. (1998). Who's afraid of a little risk? New evidence for general risk aversion. In J. Shanteau, B. A. Mellers, & D. Schum (Eds.), *Decision Research from Bayesian Approaches to Normative Systems: Reflections on the Contributions of Ward Edwards,* pp. 53-64. Norwell, MA: Kluwer.

Weber, E. U., Blais, A.-R., Betz, E. (2002). A Domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*, 263-290. doi:10.1002/bdm.414

Weber, E. U., & Hsee, C. K. (1998). Cross-cultural differences in risk perception but cross-cultural similarities in attitudes towards risk. *Management Science, 44*, 1205-1217.

Weber, E. U., & Milliman, R. (1997). Perceived risk attitudes: Relating risk perception to risky choice. *Management Science, 43*, 122-143.

Weller, J. A., Levin, I. P., Shiv, B., & Bechara, A. (2007). Neural correlates of adaptive decision making for risky gains and losses. *Psychological Science,18,* 958-964. doi:10.1111/j.1467-9280.2007.02009.x

West, R. (2008). The psychology of security. *Communications of the ACM, 51,* 34-40. doi:10.1145/1330311.1330320

Xue, G., Lu, Z., Levin, I. P., Weller, J. A., Li, X., & Bechara, A. (2009). Functional dissociations of risk and reward processing in the medial prefrontal cortex. *Cerebral Cortex, 19,* 1019-1027. doi:10.1093/cercor/bhn147

Zuckerman, M. (2007). Sensation seeking and risky behavior. Washington, DC: American Psychological Association.

Appendix A

Domain-Specific Risk-Taking (Adult) Scale – Risk-taking

Self-rating (targets and judges). For each of the following statements, please indicate the **likelihood** that you would engage in the described activity or behavior if you were to find yourself in that situation. Provide a rating from *Extremely Unlikely* to *Extremely Likely*, using the following scale:

Other-rating (acquaintances). For each of the following statements, please indicate the **likelihood** that the **acquaintance you came with today** would engage in the described activity or behavior if they were to find **themselves** in that situation. Provide a rating from *Extremely Unlikely* to *Extremely Likely*, using the following scale:

Other-ratings (judges). For each of the following statements, please indicate the **likelihood** that the **PERSON IN THE VIDEO YOU JUST WATCHED** would engage in the described activity or behavior if they were to find themselves in that situation. Provide a rating from *Extremely Unlikely* to *Extremely Likely*, using the following scale:

1	2	3	4	5	6	7
Extremely Unlikely	Moderately Unlikely	Somewhat Unlikely	Not Sure	Somewhat Likely	Moderately Likely	Extremely Likely

1. Admitting that your tastes are different from those of a friend. (S)
2. Going camping in the wilderness. (R)
3. Betting a day's income at the horse races. (F/G)
4. Investing 10% of your annual income in a moderate growth mutual fund. (F/I)
5. Drinking heavily at a social function. (H/S)
6. Taking some questionable deductions on your income tax return. (E)
7. Disagreeing with an authority figure on a major issue. (S)
8. Betting a day's income at a high-stake poker game. (F/G)
9. Having an affair with a married man/woman. (E)
10. Passing off somebody else's work as your own. (E)
11. Going down a ski run that is beyond your ability. (R)
12. Investing 5% of your annual income in a very speculative stock. (F/I)
13. Going whitewater rafting at high water in the spring. (R)
14. Betting a day's income on the outcome of a sporting event (F/G)
15. Engaging in unprotected sex. (H/S)
16. Revealing a friend's secret to someone else. (E)
17. Driving a car without wearing a seat belt. (H/S)
18. Investing 10% of your annual income in a new business venture. (F/I)
19. Taking a skydiving class. (R)
20. Riding a motorcycle without a helmet. (H/S)
21. Choosing a career that you truly enjoy over a more secure one. (S)
22. Speaking your mind about an unpopular issue in a meeting at work. (S)
23. Sunbathing without sunscreen. (H/S)
24. Bungee jumping off a tall bridge. (R)
25. Piloting a small plane. (R)

26. Walking home alone at night in an unsafe area of town. (H/S)
27. Moving to a city far away from your extended family. (S)
28. Starting a new career in your mid-thirties. (S)
29. Leaving your young children alone at home while running an errand. (E)
30. Not returning a wallet you found that contains $200. (E)

Note. E=Ethical, F=Financial, H/S=Health/Safety, R=Recreational, and S=Socia

Domain-Specific Risk-Taking (Adult) Scale – Risk Perceptions

Self-rating (targets and judges): People often see some risk in situations that contain uncertainty about what the outcome or consequences will be and for which there is the possibility of negative consequences. However, riskiness is a very personal and intuitive notion, and we are interested in **your gut level assessment of how risky** each situation or behavior is.

For each of the following statements, please indicate **how risky you perceive** each situation. Provide a rating from *Not at all Risky* to *Extremely Risky*, using the following scale:

Other-ratings (acquaintances): People often see some risk in situations that contain uncertainty about what the outcome or consequences will be and for which there is the possibility of negative consequences. However, riskiness is a very personal and intuitive notion, and we are interested in **how the acquaintance you came in with today would assess how risky** each situation or behavior is.

For each of the following statements, please indicate **how risky the acquaintance you came with perceives** each situation. Provide a rating from *Not at all Risky* to *Extremely Risky*, using the following scale:

Other-ratings (judges): People often see some risk in situations that contain uncertainty about what the outcome or consequences will be and for which there is the possibility of negative consequences. However, riskiness is a very personal and intuitive notion, and we are interested in **how the person in the video you just watched would assess how risky** each situation or behavior is.

For each of the following statements, please indicate **how risky the person in the video you just watched perceives** each situation. Provide a rating from *Not at all Risky* to *Extremely Risky*, using the following scale:

1	2	3	4	5	6	7
Not at all Risky	Slightly Risky	Somewhat Risky	Moderately Risky	Risky	Very Risky	Extremely Risky

Domain-Specific Risk-Taking (Adult) Scale – Expected Benefits

Self-rating (targets judges): For each of the following statements, please indicate **the benefits** you would obtain from each situation. Provide a rating from **1 to 7**, using the following scale:

Other-ratings (acquaintances): For each of the following statements, please indicate **the benefits the acquaintance you came in with today thinks he/she** would obtain from each situation. Provide a rating from **1 to 7**, using the following scale:

Other-ratings (judges): For each of the following statements, please indicate **the benefits the PERSON IN THE VIDEO YOU JUST WATCHED thinks he/she** would obtain from each situation. Provide a rating from **1 to 7**, using the following scale:

1	2	3	4	5	6	7
No benefits At all			Moderate Benefits			Great

Appendix B

Interview Questions

For the next part of the experiment you will be interviewed on camera about your personality characteristics, risk attitudes, and risky behaviors you have engaged in in the past. Below are the specific questions you will be asked during the interview. Please take time to read each one and write your answers to each question. This will help to prepare you for the interview process and provide you with notes to refer back to during taping.

1. Please describe your personality in general.
2. Please describe your attitudes towards risk-taking.
3. Please describe a risky behavior or activity you have done in the past. Select an activity to describe that you feel comfortable sharing, is not incriminating, and you wouldn't mind if other people knew about. Think of the kind of behavior you might talk about to a person you had recently met for the first time.

Appendix C

Examples of responses given by targets during the video interviews.

1. Please describe your personality in general (Low relevance): “I am very outgoing and like to talk a lot to people I know.”
2. Please describe your attitudes towards risk-taking (Broad relevance): “Taking risks are ok as long as no one gets hurt.”
3. Please describe a risky behavior or activity you have done in the past. Select an activity to describe that you feel comfortable sharing, is not incriminating, and you wouldn’t mind if other people knew about. Think of the kind of behavior you might talk about to a person you had recently met for the first time (High relevance): “I once raced my motorcycle down a mountain at over 100 mph in traffic.”