Use Authorization

In presenting this thesis 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 thesis 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 thesis for financial gain shall not be allowed without my written permission.

Signature _____

Date _____

MOTIVATED ACCURACY: INVESTIGATING THE EFFECT OF TASK GOALS ON NORMATIVE AND DISTINCTIVE COMPONENTS OF ACCURACY OF PERSONALITY TRAIT JUDGMENTS

by

Douglas Edward Colman

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Masters of Science in the Department of Psychology

Idaho State University

Fall 2015

To the Graduate Faculty:

The members of the committee appointed to examine the thesis of Douglas Edward Colman find it satisfactory and recommend that it be accepted.

Tera D. Letzring, Ph.D.
Committee Chair

Maria M. Wong, Ph.D.
Committee Member

Donna L. Lybecker, Ph.D.
Graduate Faculty Representative

Idaho State
UNIVERSITY

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

August 7, 2014

Douglas Colman, MBA
Stop 8112
Pocatello, ID 83209

RE: Your application dated 8/7/2014 regarding study number 4135: Motivated Accuracy: Triggering Increases in Normative and Distinctive Components

Dear Mr. Colman:

Thank you for your response to requests from a prior review of your application for the new study listed above. Your study is eligible for expedited review under FDA and DHHS (OHRP) 7. Individual or group behavior designation.

This is to confirm that your application is now fully approved. The protocol is approved through 8/7/2015.

You are granted permission to conduct your study as most recently described effective immediately. The study is subject to continuing review on or before 8/7/2015, 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; fax 208-282-4723; 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

TABLE OF CONTENTS

List of Figures

List of Tables

MOTIVATED ACCURACY: INVESTIGATING THE EFFECT OF TASK GOALS ON NORMATIVE AND DISTINCTIVE COMPONENTS OF ACCURACY OF PERSONALITY TRAIT JUDGMENTS

Thesis Abstract—Idaho State University (2015)

The current investigation aimed to replicate and extend upon the findings of Biesanz and Human (2010), who discovered instructing participants to be as accurate as possible moderated accuracy of personality judgments. Specifically, using the Social Accuracy Model (Biesanz, 2010) – a componential model that provides estimates of both normative and distinctive components of accuracy – that simple task goal increased distinctive accuracy but decreased normative accuracy. That finding was not replicated in the two current samples or across two different computational approaches to normative accuracy (nomothetic and idiographic). Expanding upon the original findings, more specific task goals (e.g., instruction of elements to which attention should be paid) did not produce reliable increases to normative or distinctive accuracy, nor did they decrease accuracy. In conclusion, utilizing a task goal when judging others likely will not hurt one's level of normative or distinctive accuracy.

Chapter I: Introduction

Unlike many species, humans do not merely wander through life reacting to a series of environmental stimuli. Rather, complex interactions of environmental influences, internal motivations, and developed patterns of cognition contribute to the non-random nature of human action (e.g., Bandura, 1978, 2006; Mischel, 1968). With this in mind, personality – individuals' typical emotional, behavioral, and thought patterns (Funder, 2001) – represents an important element to one's existence (Allport, 1921).

Personality is an important aspect of individuals, as it is predictive of many outcomes. These outcomes include health and longevity, subjective well-being, interpersonal relationships, and occupational choices (Ozer & Benet-Martinez, 2006). Additionally, data from the Mills Longitudinal Study demonstrated that three of the Big Five factors (Conscientiousness, Extraversion, and Openness) are predictive of key work-related outcomes in women over a period of 50 years (George, Helson, & John, 2011).

Given the importance of personality as a predictor of long-term consequences, there is good reason for its measurement and evaluation. Take for example the finding that women high in openness had more creative accomplishments than those lower on that trait (George et al., 2011). Using the knowledge of such outcomes can be highly beneficial in a variety of situations. For instance, creativity is a highly valuable talent in the field of marketing (Ramocki, 1994). Therefore, it would be advantageous for marketing firms to seek and hire women that are high in the Big Five trait of openness.

While the above example focuses on openness in women, there has been an abundance of research that suggests that personality traits are valid predictors of employment outcomes (Hogan, Hogan, & Roberts, 1996). As such, one might ask the question – how should one go about selecting applicants with desirable traits? A common answer is to conduct in-person interviews of the most qualified applicants. Indeed, research has found interviews to be a relatively valid method for assessing applicant personality (Barrick, Patton, & Haugland, 2000; Jackson, Peacock, & Smith, 1980; van Dam, 2003; Van Iddekinge, Raymark, & Roth, 2005).

Chapter II: Literature Review

To understand how interviews can help glean insight into the personality of applicants, it is necessary to provide a brief review of the person perception literature. As an important element of day-to-day human interaction for its many implications (Cronbach, 1955), person perception automatically begins at first sight and/or communication (Asch, 1946; Bar, Neta, & Linz, 2006; Todorov, Pakrashi, & Oosterhof, 2009; Uleman, Adil Saribay, & Gonzalez, 2008; Willis & Todorov, 2006). Zero-acquaintance (first) impressions are the foundation on which future interactions with another will be based. Each observation or interaction thereafter will serve as the basis for reasoned revision of one's first impression (Asch, 1946; Mann & Ferguson, 2015). This can be seen in the job hiring process. Applicants first introduce themselves using resumes and cover letters. Then, after narrowing the pool of applicants from these zero-acquaintance judgments, managers will further refine their impressions of applicants during interviews.

While one can appreciate the influence these zero-acquaintance impressions have on subsequent interpersonal interactions, some settings can have greater relative importance than others; for instance, a job interview. If, for example, a marketing firm was hiring but the interviewer was unable to accurately perceive openness in an applicant, a worthy candidate might be overlooked. Or, from an evolutionary standpoint (Schaller, 2008), a stranger quickly approaching while one is on a stroll in a park necessitates a quick and accurate evaluation of the person to avoid a potentially dangerous situation. Alternatively, one might be seeking a romantic relationship and

therefore wishes to judge the viability of potential partners. These real-world examples stress the importance of accurate person perception, especially at zero-acquaintance.

Accuracy Paradigm

Since initial empirical interest in person perception, the methods implemented to assess the accuracy of judgments have transformed. Early on, Cronbach (1955) provided an influential critique of key issues surrounding accuracy of personality judgment. The crux of his critique was the notion that accuracy is composed of multiple components, which should be evaluated individually rather than as one latent construct. Two of these elements identified by Cronbach (1955) were stereotype accuracy and differential accuracy.

Stereotype accuracy is the correspondence between the judge's average rating of targets (i.e., a profile reflecting the average rating for each item across all targets judged) and a normative profile (i.e., a profile reflecting the 'average' person; see Figure 1 for a visualization) (Furr & Wood, 2013). While the normative profile is typically garnered from the average profile of criterion values from a large sample of individuals, it is worth noting that said criterion can be determined in multiple ways. For instance, self-reports may represent the criterion. Alternatively, a mixture of multiple information sources such as self-reports, acquaintance-reports, behavioral observations, etc. can inform the criterion value. From this definition, one can begin to understand the importance, yet inherent complications, in accounting for this component of accuracy. Specifically, the accuracy with which one uses stereotypes to inform their impression of others depends upon one's conceptualization of the 'average' person. However, the measurement of individuals' conceptualizations is derived from *a posteriori* information. Specifically,

population estimates of the ‘average’ individual, typically garnered from separate samples or the average self-reports of the judges, are used as the stereotype accuracy criteria. Considering this, it might be sensible to garner judges’ perceptions of the ‘average’ persons’ profile and use that information as the criteria in the assessment of stereotype accuracy. Logically, this might make sense for a variety of practical reasons. For instance, one may want to understand the ‘average’ employees’ profile as construed by an interviewer, thus allowing for a better analysis of the proper utilization of stereotypes.

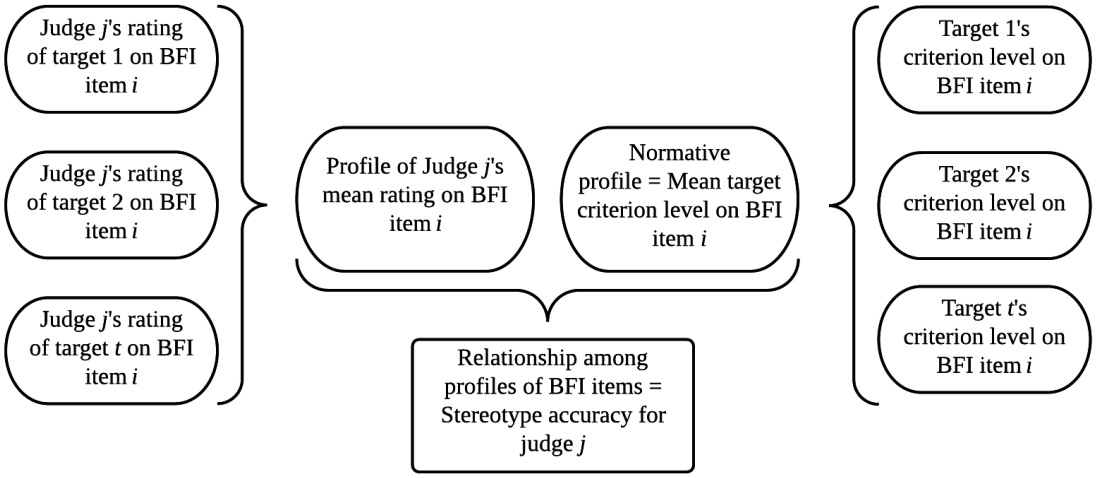

Figure 1. Visualization of stereotype accuracy computation process. BFI = Big Five Inventory.

Differential accuracy is the second notable component discussed by Cronbach (1955), and is the ability to accurately judge an individual’s level of a trait compared to the normative (‘average’ person’s) level. By extension, “[d]ifferential accuracy [also] denotes the perceiver’s ability to order the targets accurately on each trait after correcting for elevation or stereotype accuracy” (Zebrowitz, 1990, p. 81). With this in mind, differential accuracy is one of the most salient components of accuracy in the literature. However, there have been many analytic strategies employed to circumvent Cronbach’s

(1955) call for individual analysis of these components. Biesanz (2010), however, returned to the original criticism by Cronbach (1955) during the development of the Social Accuracy Model (SAM).

The SAM (Biesanz, 2010) is a componential model that allows researchers to tease apart two discrete components of accuracy in person perception. These two components of accuracy are labeled normative and distinctive in the SAM, but are equivalent to stereotype and differential accuracy, respectively, as described by Cronbach (1955). Considering the capacity of this model to resolve much of Cronbach's (1955) critique, a wealth of new research has emerged using the SAM, which is quickly becoming the preferred method of analysis within the domain of judgmental accuracy.

Realistic Accuracy Model

Given the importance of person perception, much research has been undertaken to illuminate the process by which it occurs. One such model, the Realistic Accuracy Model (RAM) as proposed by Funder (1995), describes four distinct elements – relevance, availability, detection, and utilization – which are essential to the process of person perception. First, information must be relevant to the trait being judged. If the information is not relevant, it is not useful in creating accurate impressions. Second, relevant information has to be available for perception. If information is not available, there are no valid cues on which impressions can be based. Third, relevant and available information must be detected. Even if relevant information is available, if the judge does not attend to it, accurate impressions cannot be made. And finally, the information that is detected must be utilized to generate impressions. While it is important that one detects

available relevant information, if one is unable to utilize or improperly applies that information, accurate judgments cannot, and will not, be generated.

As can be understood from the above explanations, the four elements act multiplicatively. Specifically, if one element – relevance, availability, detection, or utilization – is nil, the product of all elements become zero (Funder, 1995, p. 659). Considering that the RAM is multiplicative in nature, increases in one or more of the elements can result in meaningful changes in judgmental accuracy. Thus, research has further explicated each stage. Furthermore, there are four key moderators which "makes accuracy more or less likely" by interacting with one or more stages of the RAM – good judge, good target, good trait, and good information (Funder, 1995, p. 660).

Good judge. The first of the four moderators identified by Funder (1995) is the good judge, which asserts there to be individual differences in the ability to accurately judge others. Investigating this notion, Letzring (2008) found that several personality, behavioral, and situational factors were related to profile accuracy of personality judgment. For example, the trait of agreeableness and having 'purpose in life' were positively related to accuracy in judging another's personality.

Behaviorally, the use of basic social skills (e.g., eye contact, expressing warmth), enjoyment of the interaction, and liking of the target, among other aspects, were positively related to accuracy. Alternatively, yet intuitively, needing reassurance, undermining, and seeking advice were negatively related to accuracy. Concluding from these findings, it was asserted that such a pattern supports the idea that good judges elicit more information for use when making judgments (Letzring, 2008, p. 925).

An interesting situational factor was the impact of the number of good judges in multi-person interactions (Letzring, 2008). Because good judges are likely to elicit more relevant information (as mentioned previously), others in the multi-person interaction or even observing the interaction should also have that information available to them for making more accurate judgments. Indeed, when the composition of the group contained two good judges, greater accuracy was achieved by an observer of the recorded interaction. However, this effect was less pronounced when only one good judge was present in the group.

Viewing this moderator from another perspective, Letzring (2015) explored the relationship between beneficial outcomes and being a good judge of others' personality. Specifically, it was found that normative accuracy, or the degree to which one's judgments of others matched what the 'average' person is like, was positively related to a multitude of beneficial outcomes; such as satisfaction with life, interpersonal support and control, and positive affect. However, distinctive accuracy, or the accuracy of perceiving others as different from the normative person, was not significantly related to any of these beneficial outcomes.

Good target. The good target is the second moderator of judgmental accuracy outlined by Funder (1995). Individual differences in judgability have been found (Biesanz, 2010; Colvin, 1993), and are worth considering as a moderating variable. For instance, an individual with consistent patterns of behavior would certainly be easier to judge accurately than an inconsistent person. Similarly, one who is rather active emits more information that can be detected, thus increasing ease of judgments (Funder, 1995). Moreover, it has been found that judgable individuals are often labeled with descriptions

such as warm, cheerful, and dependable; whereas less judgable individuals are commonly labeled as hostile, deceitful, and moody (Colvin, 1993, p. 870).

A recent review identified three broad characteristic categories of this moderator – social status, psychological adjustment, and socialization – that are relevant to judgability (Human & Biesanz, 2013). The key conclusion was that being highly judgable has meaningful consequences. Specifically, these individuals tend to experience greater person-environment fit, more social support of higher quality, and less loneliness (Human & Biesanz, 2013, p. 264).

Good trait. The third moderator described by Funder (1995) is the good trait. This moderator suggests that some traits and other indices of personality are easier to judge than others. For instance, cues to highly visible traits, such as extraversion, are more likely to be detected and utilized by judges. However, there have been notable differences on which traits are considered "good" for the purpose of making accurate personality judgments. In an attempt to clarify this irregularity, Vazire (2010) suggested that good traits depend upon who is judging the trait – self, unacquainted other, or acquainted other. From this perspective, it is argued that some traits that are highly visible to oneself, like neuroticism, are not as visible to others. Thus, some traits might be accurately judged by the self, but not by others. Conversely, traits such as extraversion are commonly judged from behaviors, which are more salient to an unacquainted or acquainted other, thus allowing for more accurate judgments by others than by the self.

In a similar vein, John and Robins (1993) were interested in the evaluativeness of traits in relation to judgability. They found higher discrepancies between self-other

ratings than peer-peer ratings of traits that were highly evaluative compared to traits that were lower in evaluativeness. They ultimately reasoned that evaluative traits evoke individuals to use self-enhancement biases when making ratings. Conversely, such a biased process would not affect peer ratings as dramatically, if at all, thus resulting in discrepancies between the two rating measures. However, accuracy is not ensured by agreement among judges. Thus, such processes should not be the only evidence for assessing the accuracy of self-judgments (John & Robins, 1993).

Good information. The final of the four moderators is good information. This moderator of judgmental accuracy posits that information, over and above detection and utilization, aids in making accurate judgments (Funder, 1995, p. 660). Information quality is a powerful moderator of judgmental accuracy. To start, Andersen (1984) found that interviews based on either cognitive (thoughts) or affective (feelings) elements resulted in greater accuracy than interviews based on behavior or a mixture of these elements. From a moderator viewpoint (Funder, 1995), this could be due to thoughts and feelings revealing aspects that are considered important to the target, allowing for greater accuracy of judgments.

Along this line of thought, Letzring, Wells, and Funder (2006) investigated the effect of information quality on accuracy of judgments. Analyzing three levels of information quality, it was shown that higher quality information was positively related to accuracy of judgments while holding information quantity constant. These manipulations relate to the Andersen (1984) study in that the low-quality information condition primarily consisted of a behavioral task (completion of a trivia packet),

whereas the medium- and high-quality conditions could be said to emulate a mixed and cognitive/affective interview, respectively.

More recent research into these findings was undertaken by Letzring and Human (2014) using the SAM (Biesanz, 2010). They found that distinctive accuracy was increased while normative accuracy was decreased for select traits for judge-target dyads that engaged in discussion of thoughts and feeling or behavior as compared to engaging in activities. Moreover, it was noted that the thoughts and feelings discussion condition did not differ from the behavioral discussion condition "supporting the idea that information about thoughts and feelings is at least as useful as information about behavior" (Letzring & Human, 2014, p. 9).

Given that information quality is positively related to accuracy of judgments (Andersen, 1984; Letzring & Human, 2014; Letzring et al., 2006), Beer and Brooks (2011) examined whether different types of disclosure can alter the quality (or usefulness) of information. To do this they had participants disclose either (a) things that were important to them in life (values) or (b) personal facts that differentiated them from others (factual information). Interestingly, they discovered that accuracy of different traits was related to different disclosures. Specifically, values information was related to greater accuracy in judgments of neuroticism, whereas factual information was related to greater accuracy in judgments of conscientiousness.

As can be concluded from this discussion of the RAM and the four moderators of accurate judgment set forth by Funder (1995), a model is indispensable for generating understanding of any process and factors that may qualify it. One relevant line of inquiry is the effects of motivation, a state of being moved to do, or complete, some task (Ryan &

Deci, 2000) on interpersonal perception. Research in this domain has focused on a breadth of topics, such as empathy (Klein & Hodges, 2001), facial expression and voice tone identification (Pickett, Gardner, & Knowles, 2004), social group labels (being schizophrenic; Neuberg & Fiske, 1987), job applicant fit and potential performance (Neuberg, 1989), extraversion (Biesanz, Neuberg, Smith, Asher, & Judice, 2001) and behavior of judges related to information gathering (Biesanz et al., 2001; Neuberg, 1989).

This last topic, information gathering behaviors by judges, was considered highly important by Neuberg and colleagues (Biesanz et al., 2001; Judice & Neuberg, 1998; Neuberg, 1989; Neuberg & Fiske, 1987), as they are a driving role in the person perception process. Specifically, those motivated to create accurate impressions were more compressive in their questioning and less biased while gathering information from the target (Neuberg, 1989). However, attentional demands of the situation moderate the effect of motivation, in that distracted judges are more prone to errors due to self-fulfilling prophecies (Biesanz et al., 2001).

Within the RAM framework posited by Funder (1995), behaviors of the judge would theoretically affect the availability, detection, and utilization stages of the judgment process. As discussed earlier, behaviors of the judge are likely to elicit more trait relevant cues that are then available to the judge for detection. To this end, Neuberg and colleagues found that negative expectancies often led to self-fulfilling prophecies of those expectancies (Judice & Neuberg, 1998; Neuberg, 1989; Neuberg & Fiske, 1987). For instance, interviewers with the goal of being accurate in their impressions had less negative expectations, and thus experienced less self-fulfilling prophecies than those who sought to confirm their expectancy (Judice & Neuberg, 1998).

Although such research is important, the examination of the impact of motivation, specifically from a goal-orientation, on the accuracy of personality judgments has not yet been fully elucidated. A review of the literature returned a single investigation of such phenomenon – Biesanz and Human (2010) who investigated the role of accuracy goal orientation (i.e., accuracy-goal vs. no-goal) on the accuracy of personality judgments during initial impressions. Biesanz and Human (2010) used the SAM to capture changes in both normative and distinctive accuracy. To induce motivation, they began by first outlining the task that would be completed to both the control and experimental group participants. This task was to watch ten short videos of "getting-to-know-you" interviews after which they would complete a questionnaire about each individuals' personality. The control group participants then completed the task without any further instructions. Participants in the experimental group were given an explicit goal statement – "Since we are interested in people's personality, it is important that you form the most accurate impressions possible for each person" (p. 591). Using target self-reported personality to predict perceivers' impressions, it was found that goal-orientation (accuracy-goal vs. no-goal) significantly affected accuracy. Specifically, distinctive accuracy was greater for those with an explicit goal to be accurate. However, this increase in distinctive accuracy came at the cost of reduced normative accuracy.

Building upon the limited research on the affect of motivation on judgmental accuracy, the current study was designed to replicate and extend the findings of Biesanz and Human (2010). To do so, the accuracy-goal prompt was decomposed to also include instructions specific to normative and distinctive accuracy. Because the two types of accuracy assessed using the SAM are theoretically independent (i.e., a change in one type

of accuracy does not necessitate a change in the other; Biesanz, 2010), the increased prompt specificity aimed to reduce the 'cost' of forming a more accurate impression. Given this information, the following hypotheses were developed.

H_1: An increase in distinctive accuracy and a decrease in normative accuracy was expected to emerge for the group prompted to generate the most accurate impression possible, as compared to the control group.

H_2: An increase in normative accuracy without decrement to distinctive accuracy was expected for the group prompted to attend to the average characteristics of each individual when generating their impression, as compared to the control group.

H_3: An increase in distinctive accuracy without decrement to normative accuracy was expected for the group prompted to attend to the unique characteristics of each individual when generating their impression, as compared to the control group.

Two other aspects of the Biesanz and Human (2010) study were addressed herein. First, in contrast to only a self-rating being used as each target's criterion, a composite of a self-rating and two acquaintance-ratings (see Accuracy Criterion subsection) was implemented to ease the issues noted by Kolar, Funder, and Colvin (1996). The second feature was the use of a situation that necessitates distinctive accuracy – a job interview. A job interview provided an ecologically valid situation for the assessment of accuracy of personality judgments, as inaccuracy can have grave financial and organizational consequences (Yager, 2012).

Beyond the primary hypotheses, this study also sought to understand the implications of using a nomothetic vs. idiographic approach to assessing normative accuracy. Traditionally, a nomothetic approach has been implemented, in that a population estimate of the 'average' person's profile is used when calculating normative accuracy for all judges. In addition to the traditional nomothetic approach, an idiographic approach was also used. Specifically, instead of using a population estimate of the

'average' person's profile for each judge, each participant's perception of the 'average' person was used as the normative profile when computing normative accuracy of that judge. Because of the exploratory nature of this inquiry, no *a priori* hypotheses were devised.

Moreover, in an effort to increase generalizability of the findings, participants were sampled from two populations. Adding to the traditional in-person data collection approach, the current research utilized Amazon's Mechanical Turk (MTurk; www.mturk.com). At its core, MTurk is simply a platform for recruiting a wide array of individuals to complete what are termed "human intelligence tasks" or HITs (Mason & Suri, 2012). In a review of this new tool, Buhrmester, Kwang, and Gosling (2011) explained that researchers, or requesters as they are termed, are able to generate HITs using MTurk's provided graphic user interface or by providing a link to a traditional online data collection platform, such as SurveyGizmo or Qualtrics.

One of the major benefits of recruiting through MTurk is the low cost associated with gathering data. For a 30-minute HIT that was paying $0.02, an average of 5.3 MTurk "workers" (analogous to participants in the current writing) completed the HIT per hour. However, when compensation for the same task was increased to $0.50, an average of 16.7 HITs were completed per hour (Buhrmester et al., 2011). Such completion rates for minimal payment implies that workers are completing these tasks more out of intrinsic motivation, rather than the extrinsic reward of monetary payment (Mason & Suri, 2012).

Another key benefit of MTurk is the reduction in time required by researchers to recruit participants and collect data. As Mason and Suri (2012) explained, science

benefits from complete cycles of inquiry; that is, “from generating hypotheses to testing them, analyzing the results, and updating the theory” (p. 3). With the ability to recruit participants and collect data around the clock with MTurk, cycles can be completed at an increased pace. For this study, MTurk participants were compensated $0.50 for their efforts. Meanwhile, the traditional sample was recruited through the Idaho State University SONA research system that is managed by the Department of Psychology, and data were collected in a laboratory environment and participants were remunerated with research credits.

Chapter III: Methodology

Method

Determination of Sample Size

In determining the sample size that was necessary for the current research, conclusions by Maas and Hox (2005) for sampling in multilevel modeling were taken into account. Specifically, the authors stated that as few as 30 groups (level 2 units) is common within certain literatures (e.g., education). From simulations in which the number of groups (30, 50, & 100), group size (level 1 units; 5, 30, & 50), and the intra-class correlation (0.1, 0.2, & 0.3) were varied, it was concluded that sampling a minimum of 100 groups provided acceptably accurate regression coefficients, standard errors, and variance estimates (Maas & Hox, 2005). Extending this to the SAM (Biesanz, 2010), a cross-classified multilevel model where judge-target pairs represent the group level, including 100 judges total would satisfy the suggested minimum. However, extending beyond the principle that as few as 30 groups is common, 60 participants were recruited for each experimental condition (control, overall, and the two conditions resulting from decomposing the overall condition) within each sample (MTurk and laboratory).[1] Doing this afforded a total sample size larger than the 100 group minimum offered by Maas and Hox (2005).

Participants

A total of 589 participants were recruited from the two different sources mentioned previously and ultimately completed this study, with 298 coming from MTurk

[1] Initially, the recruitment plan was to have 40 participants per condition for each sample. However, this was increased to 60 participants per condition for each sample prior to any data being analyzed.

(64.63% Female; 79.59% Caucasian, 4.08% Hispanic, 8.16% Black/African American, 8.17% Other; M_{age} = 35.71, SD_{age} = 12.03) and 291 from Idaho State University's Psychology Department participant pool (laboratory; 72.16% Female; 77.32% Caucasian, 14.43% Hispanic, 1.03% Black/African American, 7.22% Other; M_{age} = 22.10, SD_{age} = 5.68).

Measures

Big Five inventory. The 44-item Big Five Inventory (BFI; see Appendix A; John, Naumann, & Soto, 2008) is a self-report inventory designed to measure the Big Five personality trait dimensions of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Using short phrases of basic vocabulary, personality ratings are made on a scale from 1 (*disagree strongly*) to 5 (*agree strongly*). This measure has been demonstrated to have adequate reliability with Cronbach's alpha coefficients from .79 to .88 for the five subscales and .83 for the overall measure (Benet-Martínez & John, 1998). Across both samples in this investigation, Cronbach's alpha for self-reports was .77, .80, .86, .82, and .84 for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

General demographics questionnaire. General demographics including gender, ethnicity, age, level of education, marital status, and religious affiliation were collected from participants. Further, questions related to work experience were also collected.[2] See Appendix B for the demographic and work experience information requested from participants.

[2] Ancillary to the empirical emphasis of this thesis, the Interpersonal Reactivity Index (see Appendix D; Davis, 1980), Satisfaction with Life Scale (see Appendix E; Diener, Emmons, Larsen, & Griffin, 1985), and Implicit Theory of Intelligence Measure (see Appendix F; Dweck, Chiu, & Hong, 1995) were administered.

Stimuli

Table 1

Target Demographic Information

Age (*M*)	20.83
Gender (*n*)	
Men	3
Women	3
Ethnicity (*n*)	
Caucasian	5
Black/African American	1
Relationship Status (*n*)	
Single	3
In a Relationship	2
Married	1
Religious Affiliation (*n*)	
Christian	2
LDS	1
Atheist	1
Other	2

Similar to Biesanz and Human (2010), stimulus videos were generated specifically for this project that lasted approximately 5 minutes each. Participants (N = 16) were recruited to serve as applicants for a hypothetical position in a university department office, for which the same male conducted all interviews. Six of these stimulus videos were selected on the basis of gender balance, responsiveness of the interviewees, length of interview, trait variability among the targets, and targets having two acquaintances that provided ratings (see Table 1 for demographic information of the selected targets). There were three types of interview structures that could have been implemented: structured, semi-structured, and unstructured. Structured interviews transpire in a prearranged order with no follow-up questions permitted (van der Zee, Bakker, & Bakker, 2002). Semi-structured interviews use a set of guiding questions to which follow-up questions are asked. Finally, unstructured interviews are casual, not following a predetermined set and order of questions, and applicants' answers dictate the general structure (Whiting, 2008). Noteworthy, it has been found that unstructured interviews allow for greater accuracy in judgments of personality as compared to fully structured interviews (Blackman, 2002). As such, a semi-structured interview style was chosen for the current project to promote

consistency between targets, but also allowing for relevant information about the target to become available for detection and utilization by judges. This was accomplished through the use of a list of guiding questions (see Appendix C) and subsequent follow-up questions.

Accuracy Criterion

It has been suggested that self-reports are not the most valid source to which judgments should be compared when calculating accuracy. Rather, judgments should be compared to a criterion garnered from a compilation of self-report and the reports of one's peers (Kolar et al., 1996). Following this recommendation, the current project used self-ratings and two acquaintance-ratings for each interviewee.[3] To compute the accuracy criterion, each item was averaged across the two acquaintance reports and then the resulting item mean was averaged with the self-report (see Figure 2 for a visualization of this process). The average correlation between acquaintance ratings for each interviewees' BFI profile was .50 ($p < .001$). Further, the average profile correlation between each interviewee's self-rating and the average of his/her two acquaintances ratings on the BFI was .61 ($p < .001$). The concordance between these various profile ratings provided evidence that combining these ratings was an acceptable way of establishing a single valid criterion for each BFI item for each interviewee.

Procedure

Participants were instructed to read a consent form, after which they indicated whether they would continue with the study. Next, participants were instructed to complete the self-report BFI. Additionally, participants were instructed to complete the

[3] The two acquaintances were recruited by each interviewee, with the requirement that they knew each other for a minimum of six months.

BFI about their perception of the 'average' person (see Appendix A for instructions). However, because this request had the potential to influence the ratings of the interviewees, half of the participants received this instruction after the self-report BFI while the other half received the instruction after observing and rating the stimulus videos.[4]

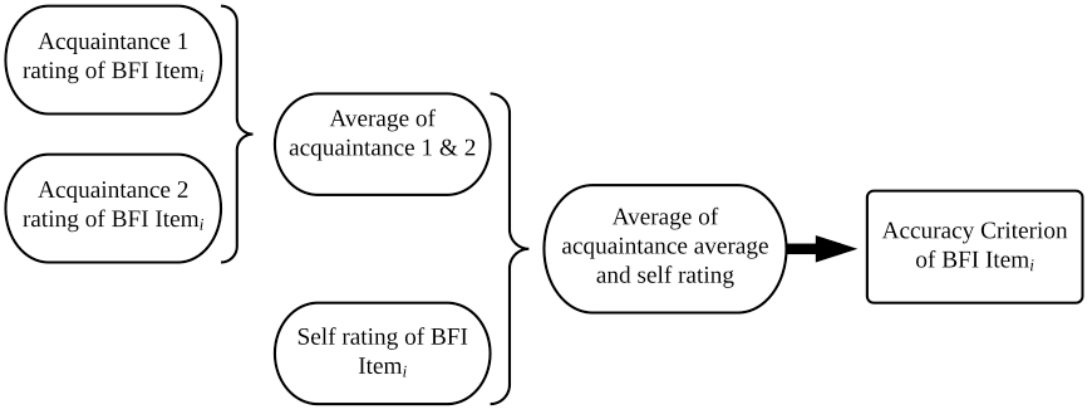

Figure 2. Visualization of the accuracy criterion computation process. BFI = Big Five Inventory.

The experimental manipulation was completed using random assignment logic in the survey tool to assign participants to one of four instruction conditions. The instructions for the first two conditions were based on Biesanz and Human (2010), while the latter two condition instructions resulted from a decomposition of the second condition to isolate normative and distinctive accuracy, respectively. In all conditions, participants read the following instructions:

> You will now view six video clips of individuals being interviewed for a secretarial position in an academic department office. After each interview you will answer a series of questions pertaining to the personality of the job applicant.

[4] This was completed using random assignment logic in the survey tool, which resulted in 50.3% (n = 296) completing the rating of the 'average' person before, and 49.7% (n = 293) completing this rating after, viewing and rating the interviewees.

In the control condition, this was the only instruction provided. For the three other conditions, this instruction was followed by additional directions related to accuracy, as follows:

> Overall motivation (OM) condition: Since we are interested in personality, it is important that you form the most accurate impression possible for each person.
>
> Normative motivation (NM) condition: Since we are interested in personality, it is important that you pay attention to how each person is normal, or how each person is similar to the average person.
>
> Distinctive motivation (DM) condition: Since we are interested in personality, it is important that you pay attention to how each person is unique, or how each person differs from the average person.

The six videos were presented in a counterbalanced order by utilizing a Latin Square design (see Table 2). After participants watched the videos and rated each interviewee using the BFI (see Appendix A for instructions), they were directed to complete the general demographics questionnaire.

Table 2

Latin Square Matrix for Video Stimulus Presentation

	Video Order					
Condition	1	2	3	4	5	6
1	E	F	C	A	D	B
2	C	E	F	B	A	D
3	A	B	E	D	F	C
4	B	D	A	F	C	E
5	D	A	B	C	E	F
6	F	C	D	E	B	A

Note. Letters A-F each represent a specific stimulus video.

Further, considering the amount of time participants spent completing study tasks, attention checks were included at multiple points within this procedure. These checks provided an indicator of the degree of adherence by participants to instructions. This was

achieved by including simple, yet easily overlooked instructions embedded within the various questionnaires (see Table 3 for the list of attention checks).

Table 3

Attention Checks Embedded Within Response Questionnaires

1. Make sure to select disagree strongly for this item
2. Make sure to select agree a little to this item
3. Do not select a response to this item
4. Select agree strongly to this item
5. Type "attention" in the text box below

Note. Participants were required to pass a minimum of 4 of these 5 attention checks.

Analytic Approach[5]

The SAM (Biesanz, 2010) was developed to simultaneously analyze both normative and distinctive accuracy. Because of the dyadic nature of the data, the SAM implements a cross-classified multilevel framework to account for the judge-target pairing (see equations 1.1 and 1.2). This is necessary as each judge can be a part of up to 6 groups (i.e., paired with each target).

$$Y_{jti} = \beta_{0jt} + \beta_{1jt}\,\mathrm{T}Crit_{ti} + \beta_{2jt}AP_i + \varepsilon_{jti} \quad (1.1)$$

$$\beta_{0jt} = \gamma_{00} + \gamma_{01}S_j + \gamma_{02}\mathrm{gender}_j + \gamma_{03}\mathrm{age}_j + u_{0j} + u_{0t} + u_{0(jt)} \quad (1.2)$$

$$\beta_{1jt} = \gamma_{10} + \gamma_{11}S_j + u_{1j} + u_{1t} + u_{1(jt)}$$

$$\beta_{2jt} = \gamma_{20} + \gamma_{21}S_j + u_{2j} + u_{2t} + u_{2(jt).}$$

Under this model, Y_{jti} is judge *j*'s rating of target *t* (the judge-target pair) on item *i* of the BFI. $\mathrm{T}Crit_{ti}$ is the criterion for target *t* on item *i* of the BFI, which was determined through self and peer ratings (see Accuracy Criterion section). AP_i is the estimate of the 'average' person on item *i* of the BFI, for which two forms were implemented. In line with the current literature, the first form was nomothetic, and was an estimate of the

[5] The lme4 package (version 1.1-10) for R (version 3.2.1) was utilized to analyze data using the SAM.

population mean for item *i* of the BFI. This estimate was gleaned by averaging the criterion scores on item *i* of the BFI across an unrelated sample of 227 targets.[6] Alternatively, the second form was idiographic, and was the unique rating by judge *j* on item *i* of the BFI for what he or she perceived as the 'average' person. It is important to note that both T*Crit*$_{ti}$ and *AP*$_i$ were mean centered prior to data analysis. The intercept of the level 1 model (β_{0jt}) is the average predicted value of judge *j*'s rating of target *t* on item *i* of the BFI when T*Crit*$_{ti}$ and *AP*$_i$ are at their mean level. In a similar vein, β_{1jt} is the average change in judge *j*'s rating of target *t* on item *i* for a one unit increase in the target's criterion value on item *i* while holding the estimate of the 'average' person on item *i* at the mean value, and is the estimate of distinctive accuracy. Likewise, β_{2jt} is the average change in judge *j*'s rating of target *t* on item *i* for a one unit increase in the estimate of the 'average' person on item *i* while holding the target's criterion on item *i* at the mean value, and is the estimate of normative accuracy.

Because the current investigation implemented two different samples (*S*; coded -1 = MTurk, 1 = laboratory), this factor was entered as a level 2 predictor in the intercept and both slope equations. Thus, two cross-level interactions were created – between *S* and T*Crit*$_{ti}$ and *AP*$_i$, respectively. Beyond this, gender (coded 0 = Female, 1 = Male) and age (mean centered) of the judge were entered as level 2 predictors of the intercept. Thus at level 2, γ_{00} represents the average intercept for an average aged female participant across samples. γ_{10} and γ_{20} represent the mean distinctive accuracy and normative

[6] This sample included 102 female and 112 male targets with age ranging from 17 to 44 (*M* = 20.56, *SD* = 3.17) and an ethnic distribution of 15.4% Caucasian, 39.6% Asian, 17.2% Mexican, 11.9% African American, 3.1% Latino, 0.4% Native American, and 6.6% Other. The criterion for this sample was calculated using the same procedure and method of acquaintance recruitment as described in the Accuracy Criterion section.

accuracy across samples, respectively. Additionally, as a random intercept and random slopes model, residual terms are included in each level 2 equation. u_{0j}, u_{1j}, and u_{2j} represent the residual variance attributed to the judge for the intercept, distinctive accuracy, and normative accuracy, respectively. u_{0t}, u_{1t}, and u_{2t} represent the residual variance attributed to the target for the intercept, distinctive accuracy, and normative accuracy, respectively. $u_{0(jt)}$, $u_{1(jt)}$, and $u_{2(jt)}$ represent the residual variance attributed to the specific judge-target combination for the intercept, distinctive accuracy, and normative accuracy, respectively.

To evaluate the two types of accuracy across groups, a series of moderation analyses were conducted using dummy coding. In the case of testing H_1 – the hypothesis that the control and OM groups would differ on both normative and distinctive accuracy – only one dummy variable ($Cond_j$; coded as 0 = control, 1 = OM) was necessary. Entering this variable into the model produced equations 2.1 and 2.2, and provided a test of the experimental manipulation. Specifically, the coefficient γ_{10} represents the average level of distinctive accuracy for the control condition. Thus, γ_{12} represents the average difference in distinctive accuracy between the control and OM conditions holding the influence of *S* constant. Finally, γ_{13} is the change in distinctive accuracy given the combined influence of *S* and $Cond_j$. Paralleling this interpretation of coefficients, γ_{20} represents the average level of normative accuracy for the control condition, γ_{22} represents the average difference in normative accuracy between these the control and OM conditions holding the influence of *S* constant, and γ_{23} is the change in normative accuracy given the combined influence of *S* and $Cond_j$.

$$Y_{jti} = \beta_{0jt} + \beta_{1jt}\,\mathrm{T}Crit_{ti} + \beta_{2jt}AP_i + \varepsilon_{jti} \quad (2.1)$$

$$\beta_{0jt} = \gamma_{00} + \gamma_{01}S_j + \gamma_{02}\mathrm{gender}_j + \gamma_{03}\mathrm{age}_j + u_{0j} + u_{0t} + u_{0(jt)} \quad (2.2)$$

$$\beta_{1jt} = \gamma_{10} + \gamma_{11}S_j + \gamma_{12}\mathrm{Cond}_j + \gamma_{13}S_j{*}\mathrm{Cond}_j + u_{1j} + u_{1t} + u_{1(jt)}$$

$$\beta_{2jt} = \gamma_{20} + \gamma_{21}S_j + \gamma_{22}\mathrm{Cond}_j + \gamma_{23}S_j{*}\mathrm{Cond}_j + u_{2j} + u_{2t} + u_{2(jt).}$$

Extending this procedure to the analysis of H_2 and H_3, three dummy variables (G1, G2, and G3) were implemented to evaluate the two types of accuracy across groups, and produced equations 3.1 and 3.2. Using this technique, a 1 was assigned for the OM group and a 0 for the control, NM, and DM groups in the first dummy variable (G1), a 1 was assigned for the NM group and a 0 for the control, OM, and DM groups in the second dummy variable (G2), and finally a 1 was assigned to the DM group and a 0 to the control, OM and NM groups for the third dummy variable (G3). See Table 4 for the dummy code matrix formed by this procedure.

$$Y_{jti} = \beta_{0jt} + \beta_{1jt}\,\mathrm{T}Crit_{ti} + \beta_{2jt}AP_i + \varepsilon_{jti} \quad (3.1)$$

$$\beta_{0jt} = \gamma_{00} + \gamma_{01}S_j + \gamma_{02}\mathrm{gender}_j + \gamma_{03}\mathrm{age}_j + u_{0j} + u_{0t} + u_{0(jt)} \quad (3.2)$$

$$\beta_{1jt} = \gamma_{10} + \gamma_{11}S_j + \gamma_{12}\mathrm{G1}_j + \gamma_{13}\mathrm{G2}_j + \gamma_{14}\mathrm{G3}_j + \gamma_{15}S_j{*}\mathrm{G1}_j + \gamma_{16}S_j{*}\mathrm{G2}_j + \gamma_{17}S_j{*}\mathrm{G3}_j + u_{1j} + u_{1t} + u_{1(jt)}$$

$$\beta_{2jt} = \gamma_{20} + \gamma_{21}S_j + \gamma_{22}\mathrm{G1}_j + \gamma_{23}\mathrm{G2}_j + \gamma_{24}\mathrm{G3}_j + \gamma_{25}S_j{*}\mathrm{G1}_j + \gamma_{26}S_j{*}\mathrm{G2}_j + \gamma_{27}S_j{*}\mathrm{G3}_j + u_{2j} + u_{2t} + u_{2(jt).}$$

While the analysis of H_2 and H_3 is more complex than for H_1, the interpretation is similar. γ_{10} still represents the average level of distinctive accuracy for the control condition. However, the extension occurs for γ_{12}, γ_{13}, and γ_{14}. These coefficients represent the average difference in distinctive accuracy between the control group and the OM, NM, and DM groups, respectively. Lastly, γ_{14}, γ_{15}, and γ_{16} are the average difference in distinctive accuracy given the combined influence of *S* and G1, G2, and G3,

respectively. Similarly, γ_{20} continues to represent the average level of normative accuracy for the control group, γ_{22}, γ_{23}, and γ_{24} represent the average difference in normative accuracy between the control group and OM, NM, and DM groups, respectively, and γ_{14}, γ_{15}, and γ_{16} are the average difference in distinctive accuracy given the combined influence of *S* and G1, G2, and G3, respectively.

Table 4

Dummy Coding Matrix Produced for Moderation Analyses

	Dummy Code		
Group	G1	G2	G3
Control	0	0	0
OM	1	0	0
NM	0	1	0
DM	0	0	1

Note. OM = Overall motivation group. NM = Normative motivation group. DM = Distinctive motivation group.

Chapter IV: Results

Results

Determination of Inclusion

The criteria for participants to be included in data analyses were chosen prior to the onset of data collection. The criteria were (A) a minimum of 80% complete data, (B) a minimum of 4 of the 6 targets were rated[7], and (C) a minimum of 80% correctness on the attention checks. This resulted in 88.6% and 86.9% inclusion for the MTurk (N = 264; 65.53% Female; 79.55% Caucasian, 3.78% Hispanic, 7.20% Black/African American, 9.47% Other; M_{age} = 35.54, SD_{age} = 12.19) and laboratory samples (N = 253; 71.94% Female; 79.45% Caucasian, 12.25% Hispanic, 0.79% Black/African American, 7.51% Other; M_{age} = 22.08, SD_{age} = 5.48). However, recent research (Hauser & Schwarz, 2015a, 2015b) suggested that invalid data does not always result from participants who fail attention check criteria. As such, the analyses for H_1, H_2, and H_3 were conducted twice – once for those who met criterion A and B, and a second time for those who met all three stipulated inclusion criterion. Because the conclusions from these two sets of analyses were different, only the results from analyses stemming from those who met all inclusion criteria are reported henceforth.

Overall Accuracy Across the Two Samples[8]

Recall that participants were recruited from two different populations – the MTurk online worker community and the Psychology Department research pool at Idaho State University. While statistical power would have been improved by simply combining them, it was necessary to test for differences across samples. This was

[7] Targets were not rated if the judge recognized them.
[8] Degrees of freedom for these models = 517 judges – 6 fixed effects – 1 = 510

accomplished by entering sample as a moderator of the two types of accuracy (normative and distinctive) in equation 1.1 and 1.2. All results will be presented for each type of normative profile.

Nomothetic normative profile.[9] Across samples, participants achieved significant levels of both normative and distinctive accuracy, $b = 0.50$, $p < .001$ and $b = 0.29$, $p < .001$, respectively. Moreover, the sample from which participants were recruited was a significant moderator of both normative and distinctive accuracy, $b = 0.05$, $p = .003$ and $b = -0.05$, $p < .001$, respectively. Interestingly, normative accuracy was greater in the laboratory sample as compared to the MTurk sample, while distinctive accuracy was greater in the MTurk sample (see Figure 3).

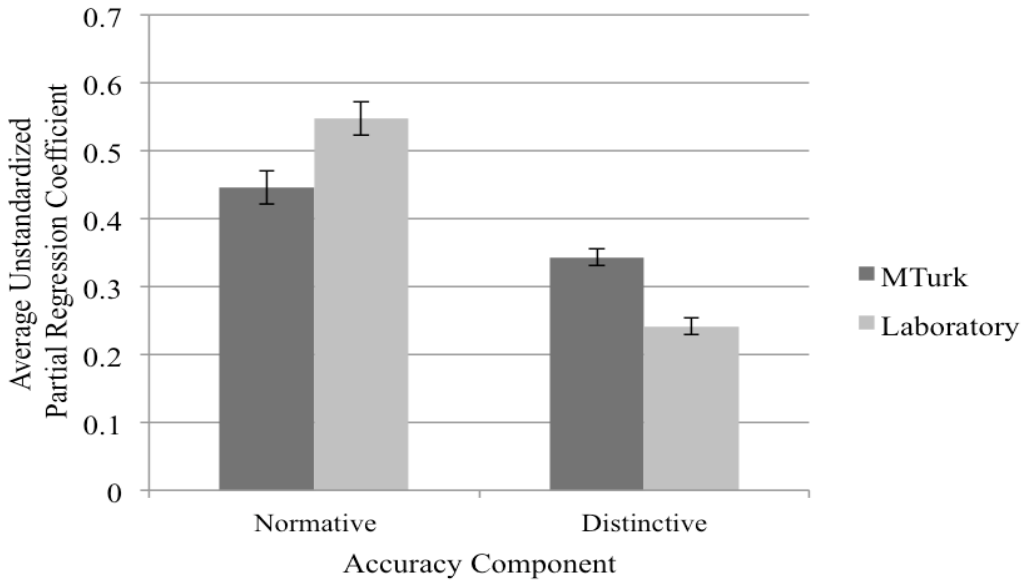

Figure 3. Results of the data for sample (MTurk and Laboratory) as a moderator using the nomothetic model is shown by plotting normative and distinctive accuracy for each sample. The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

[9] Of the two covariates tested (gender and age of the judge), only age was significant. However, the inclusion of age in all subsequent nomothetic models did not change the interpretation of the results. Thus, both gender and age were dropped from the models in favor of model simplicity.

Idiographic normative profile.[10] Across both samples, participants achieved significant levels of both normative and distinctive accuracy, $b = 0.40$, $p < .001$ and $b = 0.36$, $p < .001$, respectively. However, sample was not a significant moderator of either normative or distinctive accuracy, $b = -0.01$, $p = .42$ and $b = -0.02$, $p = .07$, respectively.

Hypothesis 1

To test H_1 – that the OM group would be lower in normative accuracy and higher in distinctive accuracy than the control group – data were subset to include only those judges in the control and OM conditions and tested using equations 2.1 and 2.2.

Nomothetic normative profile. Seeing that sample was found to be a significant moderator of both normative and distinctive accuracy across all participants, it was also important to test whether the moderation of accuracy by sample was also qualified by participant condition. This was achieved by crossing each type of accuracy with sample and the dummy variable in the model. The three-way interactions were not significant (ps > .32), which suggested that sample affected accuracy uniformly across conditions. As such, sample was retained only as a moderator of both normative and distinctive accuracy to account for variance that would otherwise be misspecified as error. In this reduced model[11], sample remained a significant moderator of accuracy with normative accuracy being greater in the laboratory sample, $b = 0.05$, $p = .009$, and distinctive accuracy greater for the MTurk sample, $b = -0.05$, $p < .001$, respectively. However,

---

[10] Of the two covariates tested (gender and age of the judge), only age was significant. With age entered in all subsequent idiographic models, only one finding changed from non-significant to significant – the interaction of sample and distinctive accuracy. However, this was not a substantive change as the significance test and estimate changed from $t(510) = -1.80$, $b = -.0173$, $SE = .0095$ to $t(497) = -2.00$, $b = -.0187$, $SE = .0096$. Furthermore, this did not change the interpretation of the results. Thus, both gender and age were dropped from the models in favor of model simplicity.

[11] Degrees of freedom for this model = 269 judges – 9 fixed effects – 1 = 259

contrary to the finding of Biesanz and Human (2010), neither normative nor distinctive accuracy was significantly moderated by experimental condition, $b = 0.08$, $p = .08$ and $b = 0.02$, $p = .42$, respectively (see Figure 4). In fact, the trend here was an increase to normative accuracy rather than a decrease.

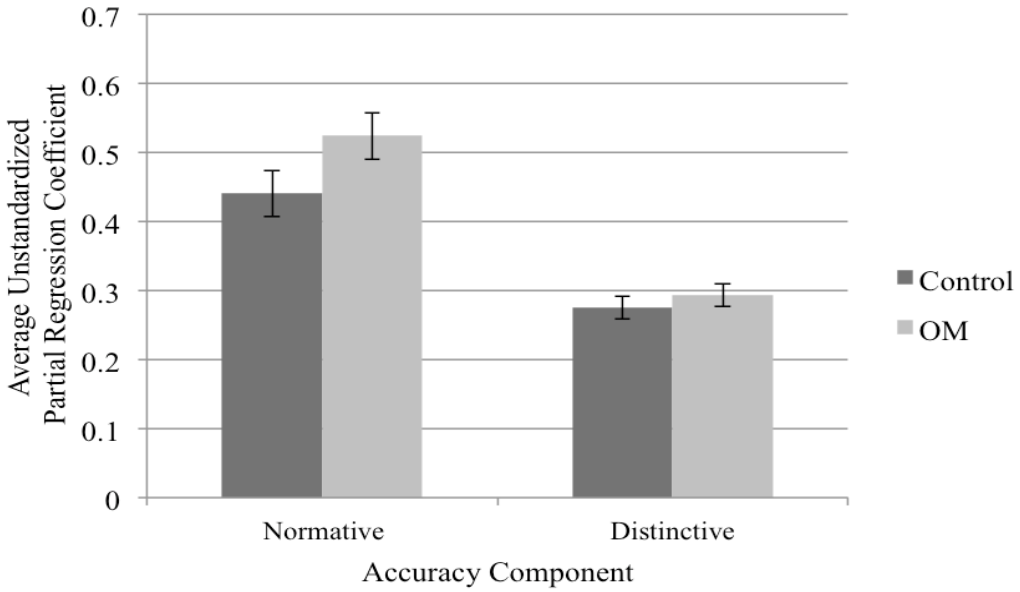

Figure 4. OM = Overall Motivation. Results of the data for Hypothesis 1 using the nomothetic model is shown by plotting normative and distinctive accuracy of the control and OM conditions averaged across the two samples (MTurk and Laboratory). The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Idiographic normative profile. Although sample was not found to be a significant moderator of both normative and distinctive accuracy across all participants, it was still important to test whether sample moderated either type of accuracy when also qualified by participant condition. This was tested by crossing each type of accuracy with sample and the dummy variable for condition in the model. The three-way interactions were not significant (*p*s > .49), which suggested that sample affected accuracy uniformly across conditions. As such, sample was completely dropped as a

moderator within the model. This reduced model[12], like the nomothetic model presented above, indicated that neither normative nor distinctive accuracy were moderated by participant condition, $b = 0.03$, $p = .29$ and $b = 0.04$, $p = .10$, respectively (see Figure 5).

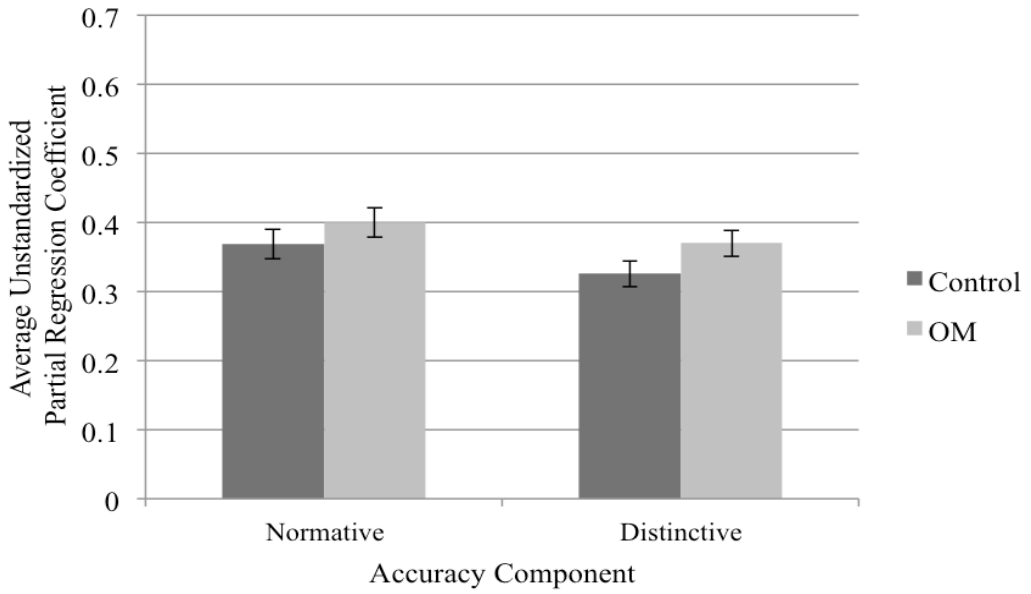

Figure 5. OM = Overall Motivation. Results of the data for Hypothesis 1 using the idiographic model is shown by plotting normative and distinctive accuracy of the control and OM conditions. The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Hypotheses 2 and 3

Equations 3.1 and 3.2 were implemented to test the next two hypotheses. The control condition served as the comparison group for both H_2 – the NM group would be higher in normative accuracy and have similar distinctive accuracy than the control group – and H_3 – the DM group would be similar in normative accuracy and have higher distinctive accuracy than the control group.

[12] Degrees of freedom for this model = 269 judges – 6 fixed effects – 1 = 262

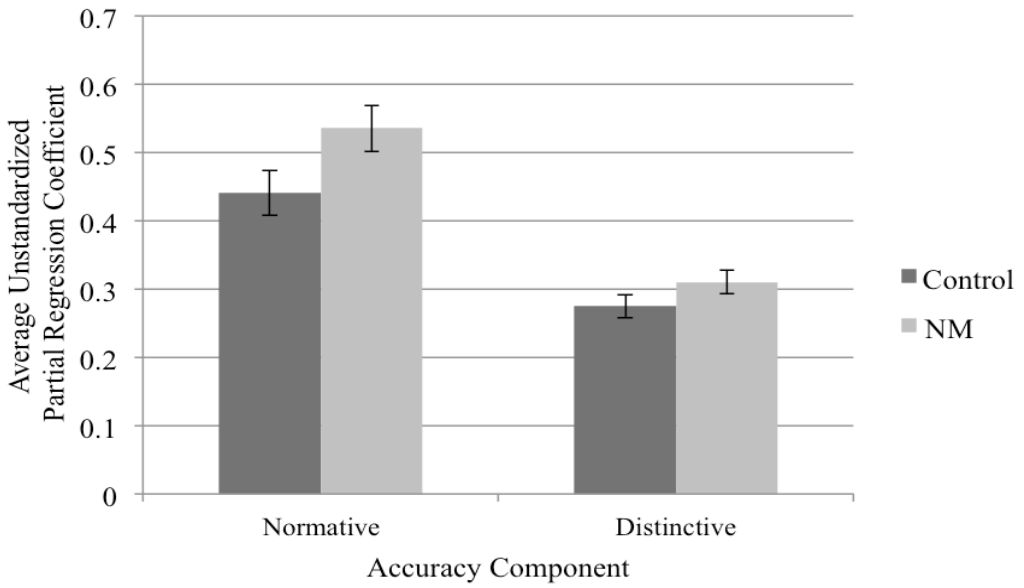

Figure 6. NM = Normative motivation. Results of the data for Hypothesis 2 using the nomothetic model is shown by plotting normative and distinctive accuracy of the control and NM conditions averaged across the two samples (MTurk and Laboratory). The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Nomothetic normative profile. Again, it was important to test whether the moderation of accuracy by sample was also qualified by participant condition. None of these three-way interactions were significant (all *p*s > .31), which suggested that sample affected accuracy consistently across conditions. As such, sample was retained as a moderator of both normative and distinctive accuracy to account for variance that would otherwise be misspecified as error. Using this reduced model[13], sample persisted as a significant moderator of accuracy with normative accuracy being greater in the laboratory sample, $b = 0.05$, $p = .003$, and distinctive accuracy greater for the MTurk sample, $b = -0.05$, $p < .001$. In full support of H_2, normative accuracy significantly increased for those in the NM group as compared to the control group, $b = 0.09$, $p = .05$, but inclusion in the NM group did not significantly alter the level of distinctive accuracy, $b = 0.04$, $p =$

[13] H_2 and H_3 were analyzed using the same model for which degrees of freedom = 517 judges – 15 fixed effects – 1 = 501

.14 (see Figure 6). Alternatively, H_3 was only partially supported in that distinctive accuracy was not significantly increased for those in the DM group compared to the control group, $b = 0.01$, $p = .58$, while as predicted, inclusion in the DM group did not significantly alter the level of normative accuracy, $b = 0.05$, $p = .35$ (see Figure 7).

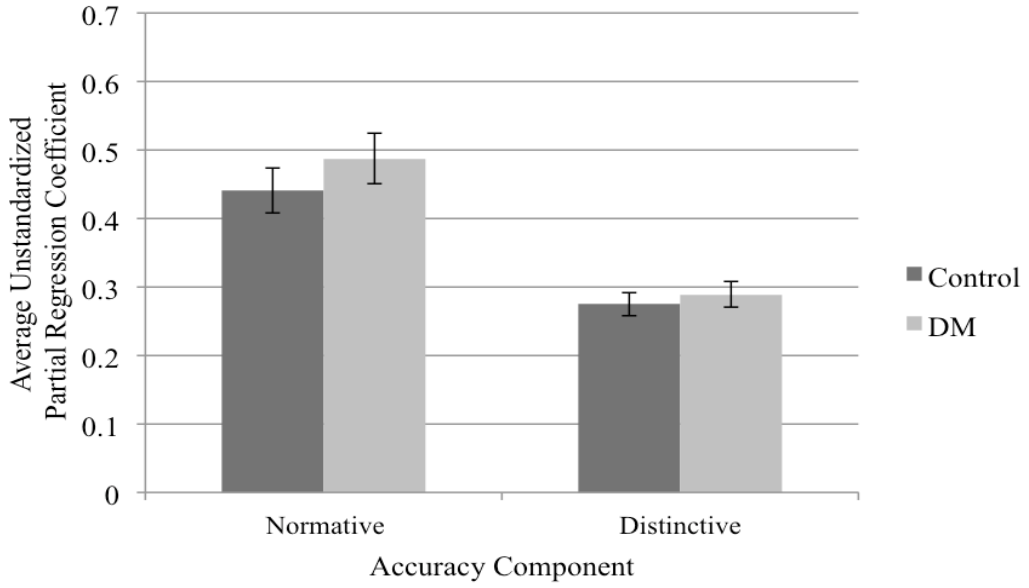

Figure 7. DM = Distinctive motivation. Results of the data for Hypothesis 3 using the nomothetic model is shown by plotting normative and distinctive accuracy of the control and DM conditions averaged across the two samples (MTurk and Laboratory). The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Idiographic normative profile. The three-way interactions between type of accuracy, sample, and experimental condition were not significant (all *p*s > .49), which indicated that sample affected accuracy equivalently across conditions. As such, sample was completely dropped as a moderator for this model. Contrary to H_2, using this model[14], inclusion in the NM group did not significantly alter the level of normative accuracy, $b = 0.06$, $p = .06$. Further, distinctive accuracy was significantly increased for those in the NM group as compared to the control group, $b = 0.06$, $p = .04$ (see Figure 8).

[14] H_2 and H_3 are analyzed using the same model for which degrees of freedom = 517 judges – 12 fixed effects – 1 = 504

Alternatively, H_3 was partially supported in that inclusion in the DM group did not significantly alter the level of normative accuracy, $b = 0.03$, $p = .37$. However, distinctive accuracy was also not significantly increased for those in the DM group compared to the control group, $b = 0.03$, $p = .35$ (see Figure 9).

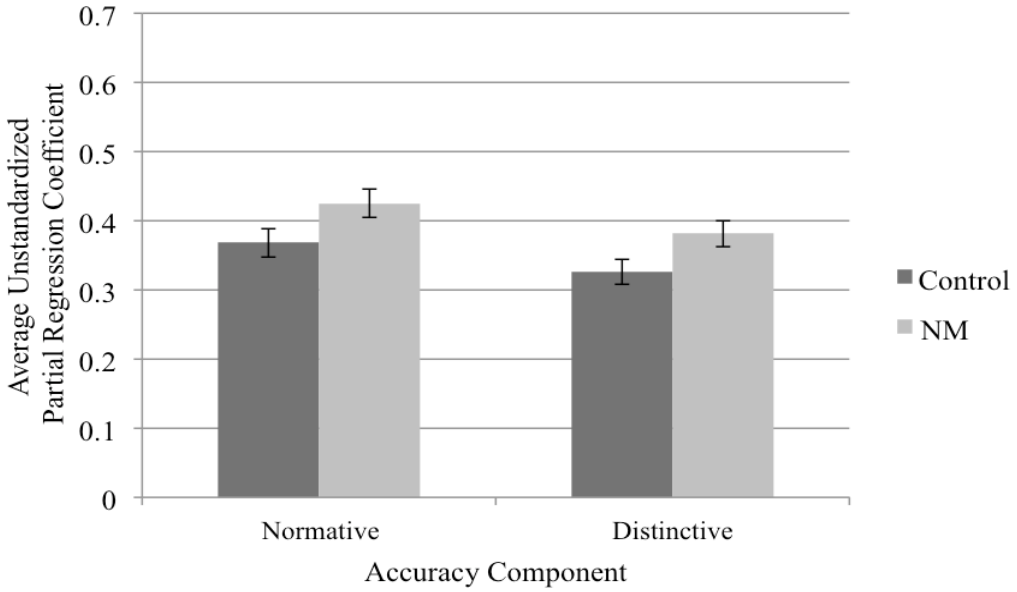

Figure 8. NM = Normative motivation. Results of the data for Hypothesis 2 using the idiographic model is shown by plotting normative and distinctive accuracy of the control and NM conditions. The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Exploratory Analyses

Moving beyond the three *a priori* hypotheses, a model comparing the NM and DM groups to the OM condition, which was identical to the accuracy motivation prompt used by Biesanz and Human (2010), was explored. Specifically, data were subset to include only those in the OM, NM, and DM groups and then analyzed using equations 4.1 and 4.2 using the OM group as the reference.

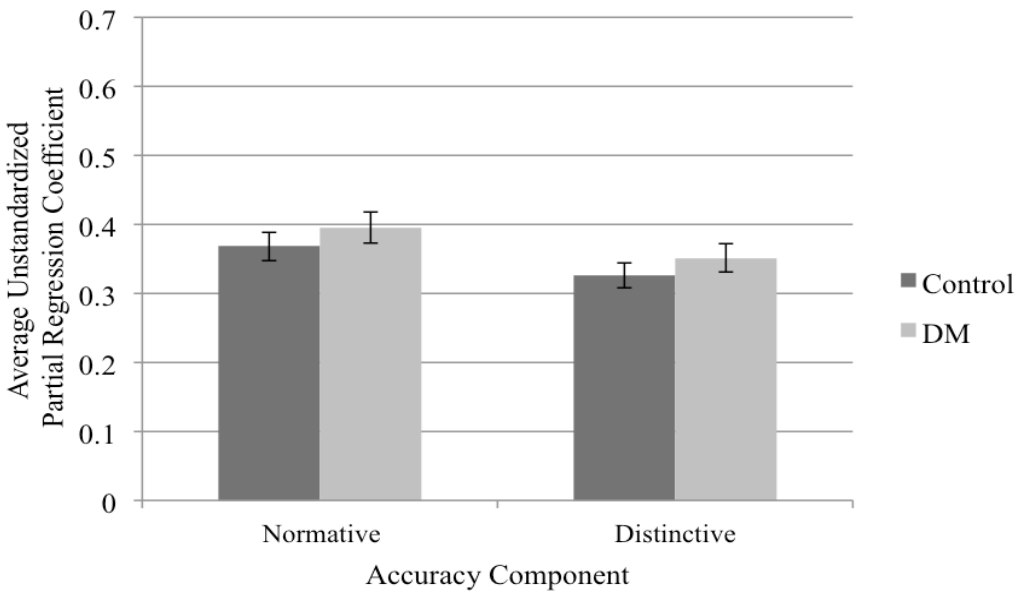

Figure 9. DM = Distinctive motivation. Results of the data for Hypothesis 3 using the idiographic model is shown by plotting normative and distinctive accuracy of the control and DM conditions. The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

$$Y_{jti} = \beta_{0jt} + \beta_{1jt}\,\mathrm{T}Crit_{ti} + \beta_{2jt}AP_i + \varepsilon_{jti} \quad (3.1)$$

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

$$\beta_{1jt} = \gamma_{10} + \gamma_{11}S_j + \gamma_{12}\mathrm{G1}_j + \gamma_{13}\mathrm{G2}_j + \gamma_{14}S_j{*}\mathrm{G1}_j + \gamma_{15}S_j{*}\mathrm{G2}_j + u_{1j} + u_{1t} + u_{1(jt)}$$

$$\beta_{2jt} = \gamma_{20} + \gamma_{21}S_j + \gamma_{22}\mathrm{G1}_j + \gamma_{23}\mathrm{G2}_j + \gamma_{24}S_j{*}\mathrm{G1}_j + \gamma_{25}S_j{*}\mathrm{G2}_j + u_{2j} + u_{2t} + u_{2(jt)}.$$

Here, γ_{10} represents the average level of distinctive accuracy for the OM group across samples. γ_{12} and γ_{13} represent the average difference in distinctive accuracy between the OM group and the group indicated by G1 and G2, respectively. Lastly, γ_{14} and γ_{15} are the average difference in distinctive accuracy given the combined influence of *S* and G1 and G2, respectively. Similarly, γ_{20} represents the average level of normative accuracy for the OM group across samples and γ_{22} and γ_{23} represent the average difference in normative accuracy between the OM group and the group indicated by G1 and G2, respectively, and γ_{24} and γ_{25} are the average difference in normative accuracy given the combined influence of *S* and G1 and G2, respectively.

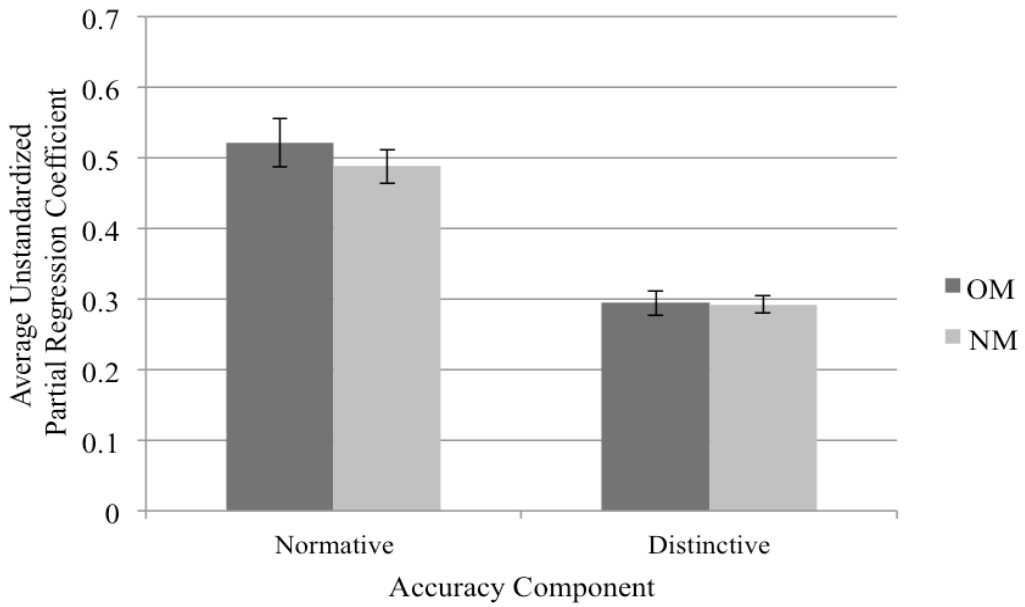

Figure 10. OM = Overall Motivation, NM = Normative motivation. Results of the data for exploratory analysis on the NM condition compared to the OM condition using the nomothetic model, shown by plotting normative and distinctive accuracy for each condition averaged across the two samples (MTurk and Laboratory). The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Nomothetic normative profile. As done previously, each type of accuracy was entered into an interaction with sample and the two dummy variables in the model to test if the moderation of accuracy by sample was qualified by participant condition. Again, the interactions were not significant (*p*s > .39). Given this, sample was retained only as a moderator of both normative and distinctive accuracy. This reduced model[15] indicated that, like all nomothetic models presented previously, sample was a significant moderator of accuracy with normative accuracy being greater in the laboratory sample, $b = 0.04$, $p = .03$, and distinctive accuracy greater for the MTurk sample, $b = -0.05$, $p < .001$. Neither the NM nor DM condition moderated normative accuracy ($b = 0.01$, $p = .77$ and $b = -0.03$, $p = .49$, respectively) or distinctive accuracy ($b = 0.02$, $p = .54$ and $b = -0.01$, $p = .84$, respectively) when compared to the OM condition (see Figures 10 and 11).

[15] Degrees of freedom for this model = 381 judges – 12 fixed effects – 1 = 368

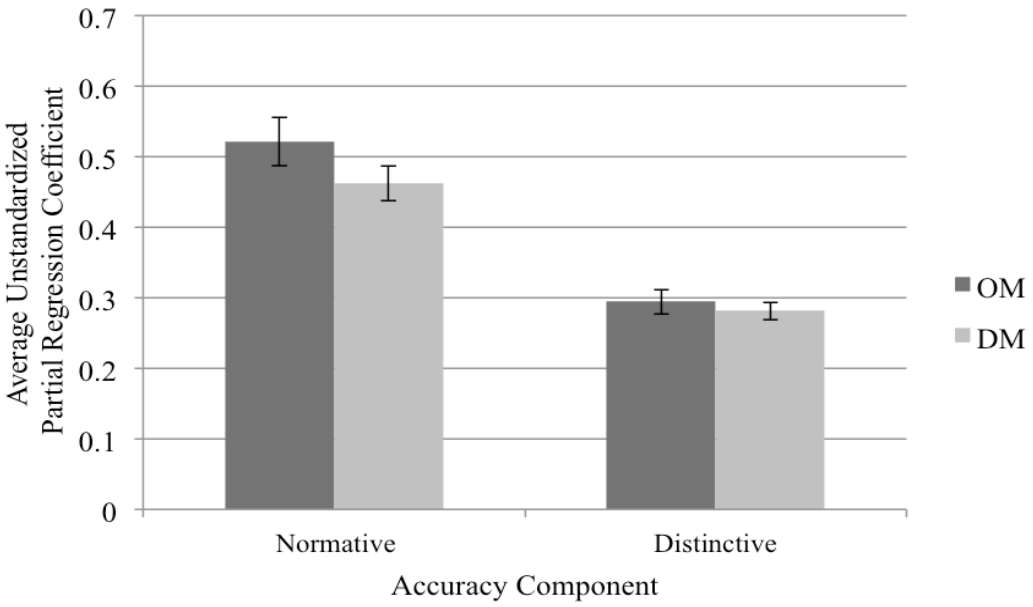

Figure 11. OM = Overall Motivation, DM = Distinctive motivation. Results of the data for exploratory analysis on the DM condition compared to the OM condition using the nomothetic model, shown by plotting normative and distinctive accuracy for each condition averaged across the two samples (MTurk and Laboratory). The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Idiographic normative profile. The three-way interactions between type of accuracy, sample, and experimental condition were not significant (*p*s > .54), which again suggested that sample affected accuracy uniformly across conditions. As such, sample was completely dropped as a moderator for this model. Ultimately, the results of this reduced model[16] paralleled those of the exploratory nomothetic model. That is, neither the NM nor DM condition moderated normative accuracy (b = 0.03, p = .41 and b = -0.01, p = .89, respectively) or distinctive accuracy (b = 0.01, p = .67 and b = -0.02, p = .51, respectively) when compared to the OM condition (see Figures 12 and 13).

[16] Degrees of freedom for this model = 381 judges – 9 fixed effects – 1 = 371

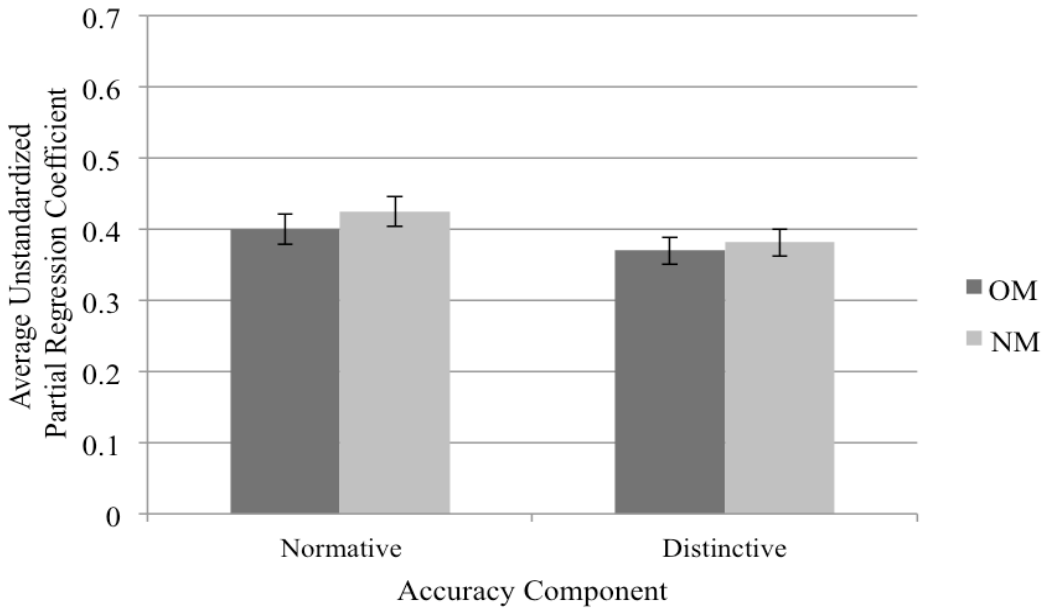

Figure 12. OM = Overall Motivation, NM = Normative motivation. Results of the data for exploratory analysis on the NM condition compared to the OM condition using the idiographic model, shown by plotting normative and distinctive accuracy for each condition. The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

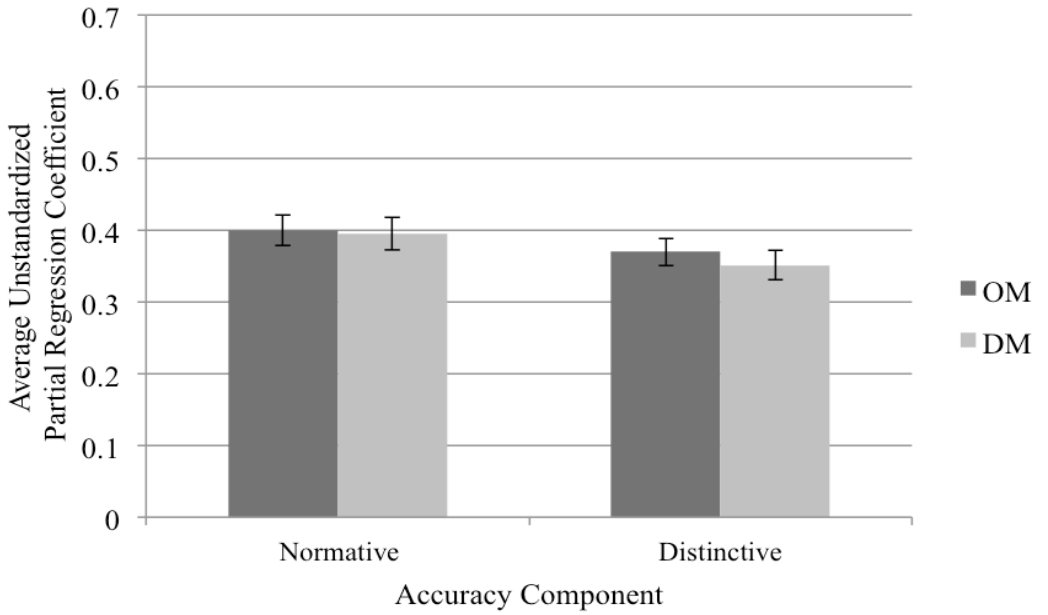

Figure 13. OM = Overall Motivation, DM = Distinctive motivation. Results of the data for exploratory analysis on the DM condition compared to the OM condition using the idiographic model, shown by plotting normative and distinctive accuracy for each condition. The graph shows the mean unstandardized fixed-effects partial regression coefficients, along with error bars representing ±1 *SE*.

Chapter V: Conclusions

General Discussion

Motivation, especially from a goals orientation, is a psychological topic that has been the subject of long standing research (see Eccles & Wigfield, 2002; Locke & Latham, 2002, 2006). However, the effect of motivation on one's accuracy in judging characteristics in others has only received scant attention. In an effort to elucidate such an effect, the current investigation was undertaken to replicate and extend one existing study (i.e., Biesanz & Human, 2010). This effort consisted of three main goals: (1) assess the effect of motivation on the accuracy of personality judgments, (2) explore the use of an idiographic approach to assessing normative accuracy, and (3) collect and analyze data from two sources in an effort to increase generalizability. There were reliable differences in accuracy of personality judgments across samples. However, taking an idiographic, rather than nomothetic, approach to the computation of normative accuracy seemed to nullify the differences in accuracy between samples. Ultimately, beyond the effect of sample, it does not seem that motivation has a robust influence on judgmental accuracy.

Effect of Task Goal

Of central importance to this thesis was the assessment of goal motivation on the accuracy of judgments of others' personalities. Because there was a limited literature on this subject, replication of a prior study was built into the design (H_1). Specifically, the same conditions that appeared in the study by Biesanz and Human (2010; control and OM) were included in this investigation. In addition to these, the OM condition was decomposed to create two new conditions with increased specificity in an effort to

increase just normative accuracy (H_2) and just distinctive accuracy (H_3), without decrement to the other type of accuracy.

Hypothesis 1, which was modeled after the findings reported by Biesanz and Human (2010), was not supported. That is, both the nomothetic and idiographic models indicated that participants in the OM group were not significantly different in their accuracy of judging targets' personality than those in the control group. Along the same line, Hypothesis 3 was only partially supported in that participants in the DM condition did not exhibit a significantly different level of normative accuracy than those in the control group. However, they did not demonstrate the expected increase in distinctive accuracy either. Furthermore, exploratory analyses assessing levels of accuracy between the OM group and the NM and DM groups were non-significant. Taken altogether, these findings suggest that a general or overall goal motivation and a distinctive goal motivation are not a reliable strategy to increasing accuracy when judging job applicants' personalities.

While Hypotheses 1 and 3 and the exploratory analyses did not yield a significant effect of a goal motivation on judgmental accuracy, Hypothesis 2 yielded an interesting mixture of results. When using the traditional nomothetic model, the hypothesis was fully supported. That is, participants in the NM condition had greater normative accuracy than those in the control condition and there was not a significant change to distinctive accuracy. However, for the idiographic approach, those in the NM group actually had greater distinctive accuracy without a significant change to normative accuracy.

Herein, the nomothetic model suggested there is an increase in the concordance between the ratings of targets' personality profiles and the estimated 'average'

personality profile for those in the NM group as compared to the control group. Conversely, the idiographic model indicated an increase in the concordance between judges' ratings of targets' criterion after partialling out their perception of the 'average' person. These two different methods of modeling normative accuracy suggest that the NM condition did not produce a reliable change in accuracy. However, it does bring to light the complexities of assessing normativeness and attempting to estimate normative accuracy (see Furr, 2008 for a discussion of such issues).

Nomothetic vs. Idiographic Approach to Normative Accuracy

One novel feature of this investigation was the use a two different approaches in the modeling of normative accuracy. The nomothetic approach has been used traditionally, and assesses the degree of correspondence between judges' ratings of targets' profiles of characteristics and an estimate of the population's profile on those characteristics (Furr, 2008). Alternatively, the idiographic approach was designed to assess the correspondence between judges' ratings of targets' profiles of characteristics and the judges' perception of the 'average' persons' profile on those characteristics.

While this is a subtle difference, the implications are meaningful. Instead of simply assessing how accurately judges perceive the normativeness of others' profiles of characteristics, one can begin to grasp how well judges implement their understanding of normativeness in that profile. This first interpretation is used by convention, and has been used as a method to partial out the effects of normativeness (e.g., typicality of characteristic profile) so that accuracy in judging targets' distinctiveness can be quantified (Furr, 2008). Alternatively, the second interpretation, and a novel contribution of this study, allows for a more individualized approach. This model allows for the

quantification of the strength of the relationship between one's perception of the generalized other and their ratings of targets. Said another way, the idiographic approach allows for exploration of stereotypes in the process of judging others.

While the utilization of the idiographic approach is a strong first step, more strides can be taken. For instance, it would be interesting to assess assumed similarity – the use of the self as a reference when making judgments (Cronbach, 1955) – in place of the traditional nomothetic approach to normative accuracy. Even further, it would be intriguing to expand the SAM to include estimates of all of these aspects – nomothetic normative accuracy, idiographic normative accuracy, and assumed similarity. Such a complex model would have several advantages. For instance, the inclusion of both the nomothetic and idiographic elements would allow the variance, typically assigned to one of these elements, to be parsed between the two. Thus, one could assess the relative predictive power of the normativeness of the profile of characteristics and one's stereotype of that profile. Furthermore, assessing assumed similarity along with both of these normativeness elements would further parse the variance to account for the use of one's self in making judgments of targets.

Sample Comparison

Data were collected from participants that were recruited from two settings – a university and MTurk. Although recruited from different locations, all participants undertook the same procedure on the Internet using a popular online survey tool. Data analyses revealed that across all nomothetic models, but not the idiographic models, participants' level of normative and distinctive accuracy were different across these two

samples. Specifically, university participants had greater normative accuracy, but lower distinctive accuracy, than MTurk participants.

This is an interesting finding on two fronts. First, it is intriguing that participants recruited from the university setting were less distinctively accurate. While it is commonly said that college populations are known to be more open to experience than those in the general community, the current sample of crowdsourced participants had higher levels of openness than the university participants (d = 0.23). Moreover, the university participants were significantly higher than the MTurk participants in extraversion (d = 0.43) and agreeableness (d = 0.29) (see also Colman, Ward, Vineyard, & Letzring, 2015). Keeping this in mind, it has been noted that openness is related to the ability to understand others and their perspectives (McCrae, 1996). Following this line of thought, it is posited that such a factor likely contributed to distinctive accuracy – the ability to differentiate targets from the 'average' person – being greater for the MTurk sample. Secondly, it is interesting to know that differences in accuracy between samples were not significant when the idiographic model was implemented. This leads back to the previous section on the subtle, but meaningful differences between the two approaches to normative accuracy. Thus, as suggested above, it would be advantageous to further explore the utility of the idiographic method to modeling normative accuracy.

Limitations

As with all research, there are limitations of this empirical investigation. To begin, the use of interviews, while ecologically valid, likely had a strong demand effect. Specifically, because the implicit and often explicit purpose of an interview is to judge an applicant, the situation likely necessitated judges' attention to available cues provided by

the applicants seen in the stimulus videos. Thus, the task demand likely reduced the effect of the provided goal motivation (Steele-Johnson, Beauregard, Hoover, & Schmidt, 2000). Given this, it would be interesting to test the conditions with increased specificity of the motivation goal using similar videos as those implemented by Biesanz and Human (2010) – of individuals answering "getting-to-know-you" questions. Such situations likely have reduced demand effects, thus allowing for greater variability in levels of accuracy.

A second, but more difficult to circumvent issue, is the use of video stimulus over live interactions. While the use of videos has proliferated in the judgmental accuracy research literature, the implementation of live, synchronous interactions is preferred to viewing contrived exchanges. Even so, the use of videotaped interactions does have some positive aspects. For instance, it allows for judges to view and rate multiple targets within a single setting, which aids in attributing accuracy affects to the judge, target, or the specific dyad. Nonetheless, it would be intriguing to have previously unacquainted participants come into a laboratory setting and get to know the other participants after being assigned to one of the four conditions presented in this investigation.

Conclusion

The results of the current study are impactful. First, the utilization of an idiographic approach to normative accuracy is a novel contribution. Prior to this study, no published research has utilized judges' perceptions of the 'average' person while assessing normative accuracy. Secondly, the discovery that when using the nomothetic approach, one should consider the source of data, as systematic differences among sampling populations can alter results and qualify claims that are made. Finally, although

nearly all hypotheses were not fully supported, it is necessary to acknowledge the fact that no significant decreases in either normative or distinctive accuracy were found. This is an important fact to keep in mind; while motivation does not seem to provide reliable increases to judgmental accuracy, it also does not seem to harm one's ability to accurately judge the personality traits of others.

References

Allport, G. W. (1921). Personality and character. *Psychological Bulletin, 18*, 441-455.

Andersen, S. M. (1984). Self-knowledge and social inference: II. The diagnosticity of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology, 46*, 294-307. doi:10.1037/0022-3514.46.2.294

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*, 258-290. doi:10.1037/h0055756

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, 344-358.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*, 164-180. doi:10.1111/J.1745-6916.2006.00011.X

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269-278. doi:10.1037/1528-3542.6.2.269

Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology, 53*, 925-951. doi:10.1111/J.1744-6570.2000.Tb02424.X

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

Benet-Martínez, V., & John, O. P. (1998). *Los Cinco Grandes* across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729-750.

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., Neuberg, S. L., Smith, D. M., Asher, T., & Judice, T. N. (2001). When accuracy-motivated perceivers fail: Limited attentional resources and the reemerging self-fulfilling prophecy. *Personality and Social Psychology Bulletin, 27*, 621-629.

Blackman, M. C. (2002). Personality judgment and the utility of the unstructured employment interview. *Basic and Applied Social Psychology, 24*, 241-250.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5. doi:10.1177/1745691610393980

Colman, D. E., Ward, D., Vineyard, J., & Letzring, T. D. (2015). Going beyond demographics: Participant characteristics to consider when crowdsourcing psychological data. *Manuscript in preperation*.

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

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin, 52*, 177-193.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85-103.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71-75.

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry, 6*, 267-285.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109-132. doi:10.1146/annurev.psych.53.100901.135153

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652-670.

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52*, 197-221. doi:10.1146/annurev.psych.52.1.197

Furr, R. M. (2008). A Framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality, 76*, 1267-1316. doi:10.1111/j.1467-6494.2008.00521.x

Furr, R. M., & Wood, D. (2013). On the similarity between exchangeable profiles: A psychometric model, analytic strategy, and empirical illustration. *Journal of Research in Personality, 47*, 233-247. doi:10.1016/j.jrp.2013.01.009

George, L. G., Helson, R., & John, O. P. (2011). The "CEO" of women's work lives: how Big Five Conscientiousness, Extraversion, and Openness predict 50 years of work experiences in a changing sociocultural context. *Journal of Personality and Social Psychology, 101*, 812-830. doi:10.1037/a0024290

Hauser, D. J., & Schwarz, N. (2015a). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavioral Research Methods*. doi:10.3758/s13428-015-0578-z

Hauser, D. J., & Schwarz, N. (2015b). It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *SAGE Open, 5*, 1-6. doi:10.1177/2158244015584617

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469-477. doi:10.1037/0003-066X.51.5.469

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

Jackson, D. N., Peacock, A. C., & Smith, J. P. (1980). Impressions of personality in the employment interview. *Journal of Personality and Social Psychology, 39,* 294-307. doi:10.1037/0022-3514.39.2.294

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift in the intergrative 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* (pp. 114-158). New York, NY: Guilford Press.

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.

Judice, T. N., & Neuberg, S. L. (1998). When interviewers desire to confirm negative expectations: Self-fulfilling prophecies and inflated applicant self-perceptions. *Basic and Applied Social Psychology, 20*, 175-190.

Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27*, 720-730.

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:10.1111/1467-6494.ep9606164112

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. (2015). Observer judgmental accuracy of personality: Benefits related to being a good (normative) judge. *Journal of Research in Personality, 54*, 51-60. doi:10.1016/j.jrp.2014.05.001

Letzring, T. D., & Human, L. J. (2014). 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

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*, 705-717. doi:10.1037/0003-066x.57.9.705

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15*, 265-268.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86-92. doi:10.1027/1614-1881.1.3.86

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology, 108*, 823-849. doi:10.1037/pspa0000021

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research Methods, 44*, 1-23. doi:10.3758/s13428-011-0124-6

McCrae, R. R. (1996). Social Consequences of experiential openness. *Psychological Bulletin, 120*, 323-337.

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

Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology, 56*, 374.

Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology, 53*, 431-444. doi:10.1037/0022-3514.53.3.431

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57,* 401-421. doi:10.1146/annurev.psych.57.102904.190127

Pickett, C. L., Gardner, W. L., & Knowles, M. (2004). Getting a cue: the need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin, 30,* 1095-1107. doi:10.1177/0146167203262085

Ramocki, S. P. (1994). It is time to teach creativity throughout the marketing curriculum. *Journal of Marketing Education, 16,* 15-25.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54-67. doi:10.1006/ceps.1999.1020

Schaller, M. (2008). Evolutionary bases of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 15-34). New York, NY: Guilford Press.

Steele-Johnson, D., Beauregard, R. S., Hoover, P. B., & Schmidt, A. M. (2000). Goal orientation and task demand effects on motivation, affect, and performance. *Journal of Applied Psychology, 85,* 724-738.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces of trustworthiness after minimal time exposure. *Social Cognition, 27,* 813-833.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology, 59,* 329-360. doi:10.1146/annurev.psych.59.103006.093707

van Dam, K. (2003). Trait perception in the employment interview: A five–factor model perspective. *International Journal of Selection & Assessment, 11*, 43-55. doi:10.1111/1468-2389.00225

van der Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology*, *87*, 176.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*, 536-552. doi:10.1037/0021-9010.90.3.536

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

Whiting, L. S. (2008). Semi-structured interviews: Guidance for novice researchers. *Nursing Standard*, *22*, 35-40.

Willis, J., & Todorov, A. (2006). First impressions making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598. doi:10.1111/j.1467-9280.2006.01750.x

Yager, F. (2012). The Cost of Bad Hiring Decisions Runs High. Retrieved from http://resources.dice.com/report/the-cost-of-bad-hiring-decisions/

Zebrowitz, L. A. (1990). *Social perception*: Thomson Brooks/Cole Publishing Co.

Appendix A

Big Five Inventory

Self-rating instructions: Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please choose a number for each statement to indicate the extent to which you agree or disagree with that statement.

Other-rating instructions: Here are a number of characteristics that may or may not apply to the person being interviewed in the video you just watched. For example, do you agree that they are someone who likes to spend time with others? Please choose a number for each statement to indicate the extent to which you agree or disagree that each statement applies to that individual.

Average person-rating instructions: Here are a number of characteristics that may or may not apply to the average person. For example, do you agree that the average person is someone who likes to spend time with others? Please choose a number for each statement to indicate the extent to which you agree or disagree that each statement applies to the average person.

1 = Disagree strongly
2 = Disagree a little
3 = Neither agree nor disagree
4 = Agree a little
5 = Agree strongly

I see myself as someone who…

1. is talkative
2. tends to find fault with others
3. does a thorough job
4. is depressed, blue
5. is original, comes up with new ideas
6. is reserved
7. is helpful and unselfish with others
8. can be somewhat careless
9. is relaxed, handles stress well
10. is curious about many different things
11. is full of energy
12. starts quarrels with others
13. is a reliable worker
14. can be tense
15. is ingenious, a deep thinker
16. generates a lot of enthusiasm
17. has a forgiving nature
18. tends to be disorganized
19. worries a lot
20. has an active imagination
21. tends to be quiet
22. is generally trusting
23. tends to be lazy
24. is emotionally stable, not easily upset
25. is inventive
26. has an assertive personality
27. can be cold and aloof
28. perseveres until the task is finished
29. can be moody
30. values artistic, aesthetic experiences
31. is sometimes shy, inhibited
32. is considerate and kind to almost everyone
33. does things efficiently
34. remains calm in tense situations
35. prefers work that is routine
36. is outgoing, sociable
37. is sometimes rude to others
38. makes plans and follows through with them
39. gets nervous easily
40. likes to reflect, play with ideas
41. has few artistic interests
42. likes to cooperate with others
43. is easily distracted
44. is sophisticated in art, music, or literature

Appendix B

Demographics Questionnaire

What is your date of birth?

- Date: _/__/__

What is your gender?

- Male
- Female
- Other? _____

What is your ethnicity?

- White
- Black/African American
- American Indian/Alaska Native
- Asian
- Hawaiian Native/Pacific Islander
- Other? _____
- More than one? _____

What is your current education level?

- Completed some high school
- High school graduate
- Completed some college
- Associate degree
- Bachelor's degree
- Completed some postgraduate
- Master's degree
- Professional degree (i.e., JD, DDS, PharmD, etc.)
- Doctoral degree

What is your marital status?

- Single (never married)
- In a relationship
- Married
- Separated
- Widowed
- Divorced
- Other? _____

What is your religious affiliation?

- Christianity
- Judaism
- Islam
- Buddhism

- Hinduism
- Spiritualism
- Agnostic
- Atheist
- Other? _____
- Prefer not to answer

At what age did you enter the workforce full-time?

- Please estimate: _____ years old

What is your current employment status?

- Full time employee
- Part time employee
- Temporary employee
- Student employee
- Non-working student
- Unemployed, but searching
- Unemployed, not searching
- Retired

How many hours per week do you USUALLY work at your job?

- Hours: _____

Counting all locations where your organization operates, what is the total number of persons employed?

- 1
- 2-9
- 10-24
- 25-99
- 100-499
- 500-999
- 1000-4,999
- 5,000+

What best describes the type of organization you work for?

- For profit
- Non-profit (religious, arts, social assistance, etc.)
- Government
- Health Care
- Education
- Other

Have you held a leadership/management position?

- Yes
- No

In total, for what period of time have you held leadership/management positions?

- Please provide an estimate: __ years and _____ months.

What level of decision-making authority do you have in your ***department***?

- Final decision-making authority (individually or as part of a group)
- Significant decision-making or influence (individually or as part of a group)
- Minimal decision-making or influence
- No input

What level of decision-making authority do you have in your ***organization***?

- Final decision-making authority (individually or as part of a group)
- Significant decision-making or influence (individually or as part of a group)
- Minimal decision-making or influence
- No input

Which of the following most closely matches your job title?

- Intern
- Entry Level
- Analyst/Associate
- Manager
- Senior Manager
- Director
- Vice President
- Senior Vice President
- C level executive (CIO, CTO, COO, CMO, etc.)
- President or CEO
- Owner

Appendix C

List of Interviewer Questions

- Describe yourself to me.
- How would your last employer describe you?
- How would your friends describe your work ethic?
- Do you get along well with coworkers?
- Do you tend to socialize with coworkers outside of work?
- How would subordinates describe you?
- Can you give me an example of when you went above and beyond in a previous job?
- Let's say that there is a spur of the moment luncheon and you have one hour to get it organized and ready… Can you walk me through your thinking and action process?
- Can you name five ways in which you could use a brick for me?
- If a supervisor was to return a task that you completed and asked you to redo it, what would it likely be and why?

Appendix D

Interpersonal Reactivity Index

Following are statements with which you may agree or disagree. For each item, indicate your agreement by selecting the appropriate number on the scale. Read each item carefully and be open and honest when answering.

1	2	3	4	5
DOES NOT DESCRIBE ME WELL				DESCRIBES ME VERY WELL

1. I daydream and fantasize, with some regularity, about things that might happen to me.
2. I often have tender, concerned feelings for people less fortunate than me.
3. I sometimes find it difficult to see things from the "other guy's" point of view.
4. Sometimes I don't feel very sorry for other people when they are having problems.
5. I really get involved with the feelings of the characters in a novel.
6. In emergency situations, I feel apprehensive and ill-at-ease.
7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.
8. I try to look at everybody's side of a disagreement before I make a decision.
9. When I see someone being taken advantage of, I feel kind of protective towards them.
10. I sometimes feel helpless when I am in the middle of a very emotional situation.
11. I sometimes try to understand my friends better by imagining how things look from their perspective.
12. Becoming extremely involved in a good book or movie is somewhat rare for me.
13. When I see someone get hurt, I tend to remain calm.
14. Other people's misfortunes do not usually disturb me a great deal.
15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
16. After seeing a play or movie, I have felt as though I were one of the characters.
17. Being in a tense emotional situation scares me.
18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
19. I am usually pretty effective in dealing with emergencies.
20. I am often quite touched by things that I see happen.
21. I believe that there are two sides to every question and try to look at them both.
22. I would describe myself as a pretty soft-hearted person.
23. When I watch a good movie, I can very easily put myself in the place of a leading character.
24. I tend to lose control during emergencies.
25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.
27. When I see someone who badly needs help in an emergency, I go to pieces.
28. Before criticizing somebody, I try to imagine how I would feel if I were in their place.

Appendix E

Satisfaction With Life Scale

Below are statements with which you may agree or disagree. For each item, indicate your agreement by selecting the appropriate number on the scale. Read each item carefully and be open and honest when answering.

1 = Strongly Disagree
2 = Disagree
3 = Slightly Disagree
4 = Neither Agree or Disagree
5 = Slightly Agree
6 = Agree
7 = Strongly Agree

1. In most ways my life is close to my ideal.
2. The conditions of my life are excellent.
3. I am satisfied with life.
4. So far I have gotten the important things I want in life.
5. If I could live my life over, I would change almost nothing.

Appendix F

Implicit Measure of Intelligence Measure

Below are statements with which you may agree or disagree. For each item, indicate your agreement by selecting the appropriate number on the scale. Read each item carefully and be open and honest when answering.

1 = Strongly agree
2 = Agree
3 = Somewhat agree
4 = Somewhat disagree
5 = Disagree
6 = Strongly disagree

1. You have a certain amount of intelligence and you really can't do much to change it.
2. Your intelligence is something about you that you can't change very much.
3. You can learn new things, but you can't really change your basic intelligence.