

ACCURACY OF GENDER AND ETHNIC LABELS AND PERSONALITY  
PERCEPTION AS MODERATED BY PREJUDICE

By

BRANDON P. LABBREE

A thesis submitted to the

Graduate School-Camden

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of Master of Arts

Graduate Program in Psychology

Written under the direction of

Wayne Chan

And approved by

---

Wayne Chan

---

Kristin August

---

Christopher Nave

Camden, New Jersey

October 2016

## ABSTRACT OF THE THESIS

Accuracy of Gender and Ethnic Labels and Personality Perception as Moderated By

Prejudice

By BRANDON P. LABBREE

Thesis Director:

Wayne Chan

Prejudice has been of great interest to the field of psychology since near its creation, but little research has been done to examine the role of individuals' prejudice on the accuracy of their judgment of others. In this study, participants read eight short self-descriptions written by undergraduate students in a previous study. The self-descriptions were labeled as having been written by undergraduate students belonging to different gender and ethnic groups. Participants then rated the personality of the authors on a series of traits. Each participant was shown two self-descriptions that were accurate in both their gender labels and ethnicity labels, two self-descriptions that were labeled with accurate gender and inaccurate ethnicity labels, two self-descriptions that were labeled with inaccurate gender labels and accurate ethnicity labels, and two self-descriptions with inaccurate gender labels and inaccurate ethnicity labels. The participants completed self-report measures of gender, racial, and general prejudice levels, as well as self-rated personality. Accuracy was assessed through the strength of intraclass correlations (ICC) between the participants' ratings and the self-reported profiles of the authors. Upon analysis of the data, the author failed to find support for the hypotheses. There appears to be no effect of

having accurate gender labels or accurate ethnicity labels on personality ratings.

Prejudice was not found to moderate this relationship.

## Table of Contents

1. Title Page	i
2. Abstract	ii
3. Table of Contents	iv
4. List of Table	v
5. List of Figures	vi
6. Introduction	1
7. Preliminary Study	18
8. Methods	21
9. Results	37
10. Discussion	46
11. Appendices	81
12. References	104

## List of Tables

Table 1: Mean Level of Rater-Target Agreement by Self-Description	55
Table 2: Correlation Matrix of Prejudice Measures	56
Table 3: Means of Rater-Target Agreement for Separate Self-Descriptions	57
Table 4: Means of Rater-Target Agreement for Self-Descriptions Collapsed	61
Table 5: Three-Way ANOVA for Neuroticism	65
Table 6: Three-Way ANOVA for Extraversion	66
Table 7: Three-Way ANOVA for Openness	67
Table 8: Three-Way ANOVA for Agreeableness	68
Table 9: Three-Way ANOVA for Conscientiousness	69
Table 10: Correlation Matrix Between Rater Personality and Rater-Target Agreement for Target Personality	70
Table 11: Correlation Matrix between Rater Personality and Rater-Target Agreement Across Targets, Controlling for Prejudice	71
Table 12: Correlation Matrix of Between Rater Personality and General Rater-Target Agreement	72
Table 13: Correlation Matrix of Between Rater Personality and General Rater-Target Agreement Controlling for Prejudice	73

## List of Figures

Figure 1: Effects of racial/ethnic label and gender label accuracy on mean rater-target agreement for self-description H	74
Figure 2: Effects of racial/ethnic label and gender label accuracy on mean rater-target agreement for self-description A	75
Figure 3: Effects of racial/ethnic label and gender label accuracy on mean rater-target agreement for self-description C	76
Figure 4: A comparison of the effects of racial/ethnic label accuracy and gender label accuracy on mean rater-target agreement by different self-descriptions.	77
Figure 5: Relationship between standardized index of global prejudice and rater-target agreement with fit line.	78
Figure 6: Relationship between standardized index of sexism and rater-target agreement with fit line.	79
Figure 7: Relationship between standardized index of racism and rater-target agreement with fit line.	80

## Introduction

### Overview of Person Perception

At a given moment, a single individual may find himself or herself interacting with individuals who differ in experiences, traits, thoughts, beliefs, emotional states, and in countless other ways. As a result, in order to navigate this complex social world the ability to understand and quickly make judgments about others is among the most important skills that humans have (Beer & Watson, 2008; Lewis, Hodges, Laurent, Srivastava, & Biancarosa, 2012; Zaki & Ochsner, 2011). In fact, it often happens almost instinctively; people categorize individuals mere moments after meeting them, using only minimal information (Letzring, Wells, & Funder, 2006). Being of relevance to how people interact with one another, and across groups, the field of social psychology has been exploring the processes of *person perception*, as the creation of judgments is called, since nearly its inception (Zaki & Ochsner, 2011).

Because of the immediate nature of these judgments, much research has been done on the knowledge required to achieve accurate impressions of others. Accurate judgments of others can be formed based on observing less than a minute of behavior (Ambady & Rosenthal, 1992; Slepian, Bogart & Ambady, 2014). In fact, this discovery of how much can be gleaned from a stranger at zero acquaintance has profoundly impacted psychological research. Research on “thin-slicing,” or the perception of others based on limited information (Ambady, 2010; Ambady & Rosenthal, 1992; Slepian, Bogart & Ambady, 2014), has been used extensively by those interested in social judgment to examine the extent that these first impressions can have a degree of accuracy above chance. The types and amount of information necessary to make such impressions

are often outside of an individual's conscious awareness (Oltmanns, Friedman, Fiedler & Turkheimer, 2004; Slepian et al., 2014). Different researchers have examined the use of minimal information such as photographs (Rule & Ambady, 2011), short audio recordings (Ambady & Rosenthal, 1992), and video recorded examples of nonverbal behavior (Slepian et al., 2014) in making judgment on strangers' levels of different traits and emotional states. Similar results have been found in recent years with the examination of digital behavior, or actions that exists either exclusively or significantly online (Gosling & Mason, 2015; Tskhay & Rule, 2014). Researchers have found that accurate judgments can be made in such domains as email addresses (Back, Schmukle & Egloff, 2008), personal webpages (Vazire & Gosling, 2004), and blog posts (Li & Chignell, 2010).

The present study builds upon this literature by examining the accuracy of judgments of individuals based upon short self-descriptions and how they may be influenced by the perceived social group membership of the authors. The perceived social group memberships of the targets may, or may not, reflect the targets' self-reported memberships. In addition, the present study will benefit the literature by seeing if a match between rater perceived social group membership and target-reported group membership will impact the accuracy of a rater's perception of a stranger based upon a self-description.

### **Personality Judgment**

One of the other fields of psychology besides social psychology that has taken the greatest advantage of research in person perception is its sister field of personality. This adoption of person perception as a major focus in the field is the result of two



interconnected movements in personality psychology. The first is a movement towards more behavioral and other external measures of personality (Baumeister, Vohs, & Funder, 2007; Funder, 2001; Furr, 2009). This has been adopted in response to some of the criticism of personality research expressed in Mischel (1968), that personality does not do a good job of predicting behavior across situations (Funder, 2009). While self-report is still considered one of the methods with which personality is best assessed, by utilizing behavioral measures for personality, some personality theorists are emphasizing the importance of behavior when discussing personality, as it is among the strongest measures of validity for personality psychology, as well as psychological science in general (Funder, 1991; Furr, 2009). For example, Tskhay and Rule (2014)'s meta-analysis demonstrates that personality can be leaked through one's written behavior to a great extent, both in actual word use as well as personality seen by raters. The existence of possible personality leakage through written behavior impacted the decision to use written self-descriptions in the current study as targets for which to have participants rate in terms of personality, as it allows the researchers to understand that the results are not the result of strictly shared stereotypes, but of actual perceptions of the targets' personalities.

With an increased emphasis on using other assessments in addition to self-reports, the field has to deal with a new set of both empirical and methodological issues. While self-reports of personality have been consistently found to be a predictor of a range of outcomes and accurate to an impressive degree, peer or observer judgments of personality are on average no less accurate (Carlson, Vazire, & Oltmanns, 2013; Human & Biesanz, 2011; Shrauger & Osberg, 1981; Vazire, 2010) and can provide an alternative

method in which to examine the psyche of the individual (Vazire, 2010; Vazire & Carlson, 2011). Observer judgments of personality, however, vary greatly in quality. For example, peer or observer reports do well when judging more external traits, such as extraversion, but do more poorly when examining more internal traits (Vazire & Carlson, 2011). At the same time, these judgments are influenced by external factors not connected with the target, such as differences in status between judges and targets (Hall, Schmid Mast, & Latu, 2015), the type of relationship that the observer has with the target (Letzring, 2015), group memberships (Rogers & Biesanz, 2014) and motivation of the observer to be accurate in their judgments (Biesanz & Human, 2010). The current study will examine how one such external factor, perceived group membership of targets, impacts the accuracy of personality judgments.

In an attempt to make the validity more comparable between self-reports and peer or observer reports, research has focused on determining what makes a good, or bad, judge of personality, in order to maximize both the consistency and the accuracy of the information that one receives from informants and other personality judges (Kenny, 1991; Letzring, 2008; McLarney-Vesotski, Bernieri & Rempala, 2011). To this effect, researchers have developed various models of personality judgments, such as Funder (1995)'s Realistic Accuracy Model, Kenny (1991)'s Weighted Accuracy Model, and Biesanz (2010)'s Social Accuracy Model in order to understand what goes into creating an accurate impression of an individual.

According to the Realistic Accuracy Model (RAM), the creation of an accurate personality judgment is the result of four independent stages: target displaying cues and behaviors relevant to the attribute being assessed, the judge observes the available cues,

the judge detects the cues, and the cues are used for the judgment (Funder, 1995, 2012). A person makes an accurate judgment when they are able to see and use relevant behavior to assess the appropriate trait. This model is enhanced with the presence of what are referred to as “good information” and “good judges”. RAM refers to good information as being in both quantity (that more information is better than less) and quality (highly relevant and diagnostic for the associated trait independent of any situational cues that limit the expression of the associated behavior). For example, people are able to make more accurate assessment of individuals’ personality in more unstructured versus structured conversations. This access to “good information” is further augmented in the presence of “good judges”, who are those individuals that have either personal expertise or ability to utilize the cues most effectively (Funder, 1995, 2012).

RAM provides a good basis by which one can understand the personality judgment process. However, it is limited by its focus on single traits. In everyday life, it is not very common for people to be interested in only one trait or characteristic of an individual. Additionally, the evidence that even thin-slice rating of strangers display some level of accuracy is not completely explained by this model. (Ambady, 2010; Ambady & Rosenthal, 1992; Carney, Colvin, Hall, 2007; Slepian et al., 2014). According to the thin-slice literature, judgments based upon less than 30 seconds of observation display a greater than chance level of accuracy. RAM would propose that this accuracy is achieved through humans being relatively good judges or provide enough “good information”, although, this explanation is weakened by the decreasing level ratio between slice length and accuracy beyond a certain point (Carney et al., 2007). As a

result, while this model provides a useful basic structure, it does not adequately take into account the breadth of information available to any rater.

The role of the “good judge” and “good information” reflects something similar to what is described by the Weighted Accuracy Model (WAM), a fact that RAM acknowledges (Funder, 1995). WAM is primarily designed to describe the how consensus is reached between multiple observers of a target; however it is a model that could logically be applied to explain self-other agreement, if one thinks of it as being a “consensus” between two judges, with the self being just a type of judge (Kenny, 1991). According to this model, judges take into account a wide range of different information that they have in order to make their judgments. The individual judge must then consider the relative weights that the different pieces hold in the decision due to their relevance and how diagnostic it is for the characteristic being judged in order to create the most accurate assessment (Kenny, 1991). While this model does account for the use of extraneous or non-behavior relevant information better than RAM, RAM would likely suggest that such information is simply not “good” information that a judge may or may not use in their judgment.

WAM includes the assumption that this information is used, and thus may influences judgments in differing amounts. However, this model places a great deal of emphasis on the interaction between the judges and how they may influence each other, as well as their differential reactions to different observed acts. In the case of self-other agreement, this interaction between judges is much more limit, and as thus its influence may not necessarily be as strong as in multiple judge scenarios, especially in scenarios where the target and non-target judge do not meet and judgments are being made after

the fact. This could cause information to flow only one way, limiting the capability to obtain consensus. Since consensus is often used as a proxy for accuracy, limiting the ability to obtain some form of consensus limited the ability to make accurate assessments of a target. As a result, this model is limited in its ability to truly explain accuracy to very specific circumstances (Kenny, 1991). In particular, the current study uses rater-target agreement as a measure of accuracy, with the targets and raters not interacting with one another. Since this creates an only one-way flow of information between “judges” (i.e. the rater and target), WAM would be of limited utility to explain the process underlying personality perception in the particular case. Like WAM, the Social Accuracy Model (SAM) assumes that the individual judgments of personality consist of multiple complex components; however, this model goes one step further by examining a target’s personality as a whole, instead of just a single trait (Rogers & Biesanz, 2014). In addition, SAM examines what it calls “impressionistic accuracy”, or the accuracy of a perceiver’s judgment of a target’s personality (Biesanz, 2010). Impressionistic accuracy is an analogous concept to the accuracy explained in RAM and WAM. Similar to RAM’s concepts of the “good target” and the “good judge”, impressionistic accuracy is composed of two subtypes: “expressive accuracy”, which is how accurate an individual is assessed across judges, and “perceptive accuracy”, how accurate a judge is able to assess targets (Biesanz, 2010).

Several elements are examined while studying impressionistic accuracy, which includes how an individual compares to the general population or another reference group. As a direct result, group membership, or at least perceived group membership plays a role in the formation of personality judgments, according to this model (Rogers &

Biesanz, 2014). It is this model that frames the work that this study aims to accomplish, as it explicitly allows for the integration of target group membership directly in judges' perceptions of target profiles of traits. Both RAM and WAM were designed more for the examination of individual traits than the entire profile produced during a judgment. SAM, on the other hand, focuses more on understanding the overall perception that a judge has for a target, and thus will be better able to explain personality perception in daily life. By splitting impressionistic accuracy into the expressive accuracy and perceptive accuracy subtypes, it also allows for further examination of the processes of personality perception and how they may work together, as well as work separately, to produce an accurate judgment. The consideration of these two subtypes as forms of accuracy in and of themselves, it allows for the field to frame the process of personality judgment as not just a single process that can be influenced by multiple factors, but rather as two interrelated processes that each has influences that result in making an accurate judgment. The advantage of thinking about personality assessment in terms of expressive and perceptive accuracies is that it suggests a little more malleability to either a target being accurately perceived across judges or a judge being accurate across multiple targets than saying an individual is a "good target" or a "good judge", which would be the language that RAM and WAM would use to describe the phenomena.

Regardless of one's preferred model of personality assessment, they are all relevant in understanding personality assessment using what has been called "behavioral residue", which is the traces of one's behavior that one leaves behind as the result of one's actions in a given environment, which can be either physical or digital (Gosling, Ko, Mannarelli, & Morris, 2002; Vazire & Gosling 2004). One type of behavior that has

the capability to leave behind behavioral residue, particularly in a digital environment, is verbal behavior; a lot can be learned about a person through the language that he or she uses. One such environment affected by observable behavior residue is an individual's verbal behavior, whether in person or through a digital medium (Pennebaker, Mehl, & Niederhoffer, 2003). For example, much work has been done to assess the personality of individuals using direct computerized analysis of the words that they used in tweets (Qiu, Lin, Ramsay, & Yang, 2012), self-narratives (Hirsh & Peterson, 2009), social media sites (Tskhay & Rule, 2014), and even text messaging (Holtgraves, 2011), finding that each medium provides the insight into the personality of its author. In addition, similar results about being able to accurately determine personality through written behavior have been found through the use of independent raters (Back et al., 2008; Gill, Oberlander, & Austin, 2006; Tskhay & Rule, 2014; Qiu et al., 2012). The current study seeks to advance the field concerning behavioral residue and its ability to accurately communicate an individual's personality by examining what role that prejudice may play in the personality assessments made by raters in such scenarios.

### **Stereotypes**

When making judgments about strangers, people are placed at a disadvantage in comparison to when they are making similar judgments of individuals that they know very well, due to a lack of information available. To combat this lack of *individuating information*, or information about the specific individual being assessed, those making judgments of strangers often have to rely on the use of *stereotypes* as a source of information about the targets (Crawford, Jussim, Madon, Cain & Stevens, 2011). Stereotypes are judgments that an individual, or individuals, have concerning a group or

members of said group (Jussim, Harber, Crawford, Cain, & Cohen, 2005). As about individuals and groups, they serve as mental shortcuts in the categorization and perception of others. For example, Darley and Gross (1983) had participants at Princeton University observe a short video recording of a young girl taking a test, but some participants were told that she was from a middle-class, suburban area, while others were told that she was from a low income, urban area. Depending upon what the participant was told, the participant rated the young girl differently on her academic skill and potential (Darley & Gross, 1983). Every participant witnessed the same young girl perform the same behavior, but yet still judged her differently. They started with preconceived notions about her due to the stereotypes of people from different socioeconomic statuses, and they perceived her behavior in reference to the particular stereotypes. This interest in stereotypes can be traced back to the early days of the field (Allport, 1954/1979; Judd & Park, 1993; Jussim, 2012; Kenny & Albright, 1987; Ottati & Lee, 1995; Schneider, 2004).

Stereotypes have generally been viewed as negative and inherently false by the field of psychology for much of the last century since they began being studied (Allport, 1954/1979; Jussim, 2012; Schneider, 2004). However, recent research has begun to question the assumption that stereotypes are inherently false (Jussim, McCauley, & Lee, 1995; Jussim, 2012; Krueger, 2012). Some researchers have gone as far as to claim that “stereotypes have been stereotyped” (Jussim, et al, 1995, p. 15). Although they suggest that it is an imperfect model, researchers argue that some stereotypes exist due to having “a kernel of truth”, meaning that these stereotypes have at least some basis in reality but are exaggerated or otherwise distorted. As a result, these stereotypes have the potential to



have at least some utility in the judgment of persons (Jussim, 2012; McCauley, 1995). Such a claim harkens back to the early days of social perception research, bringing the field full circle (Jussim, et al., 1995; Ottati & Lee, 1995). What little research on the accuracy of stereotypes that can be found, at least in terms of personality, primarily exists in cross-cultural comparisons of gender (Löckenhoff, et al., 2014), age (Chan et al., 2012), and national character stereotypes (McCrae et al., 2013). However, some other examinations of the accuracy of stereotypes do exist, such as in stereotypes of political attitudes and beliefs (Diekman, Eagly, & Kulesa, 2002) and internal thoughts (Lewis et al., 2012). These examples of accuracy can be attributed to two possible mechanisms: *social projection*, which would mean that the raters would rate the targets differently as a result of the stereotype of their group, or *self-stereotyping*, which would mean the target himself or herself would have internalized the individual stereotypes and behave accordingly (Cho & Knowles, 2013). While understanding the exact mechanism by which stereotypes may be used to produce accurate judgments of a stranger's personality is beyond the scope of this study, the methodology of the current study relies on participants utilizing social projection in their assessment of targets. This would provide some variability in the perception of a target, depending upon the individual stereotypes that are being manipulated.

### **Racial/Ethnic Stereotypes**

While just about any form of social categorization that can create groups can produce stereotypes, much of the literature has focused on primarily two types: racial/ethnic stereotypes and gender-based stereotypes. Racial/ethnic stereotypes were among the first studied by researchers, having been arguably around for centuries

(Dovidio, Gaertner & Kawakami, 2010; Gilbert, 1951; Karlins, Coffman, & Walters, 1969; Katz & Braly, 1933; Madon, Gyll, Aboufadel, Montiel, Smith, Palumbo, & Jussim, 2001). Just as with other types of stereotypes, racial/ethnic stereotypes assume that members of a given race and/or ethnic group often share, or have a greater likelihood to share than the general population, a series of characteristics, beliefs, or attitudes with one another (Dovidio et al., 2010; Terracciano et al., 2005). Understanding the use of stereotypes based on racial or ethnic group membership can be of great importance in a multicultural society such as our own, as cross-group interactions have increased greatly in recent decades (Dovidio et al., 2010). While understanding stereotypes is an important endeavor in such a climate, great care must be taken when examining these judgments; stereotypes could be reflective of the actual traits of the population or a complete falsehood, making reliance on them prone to misuse and abuse. In fact, the research on whether racial/ethnic stereotypes have any accuracy to them has been mixed (Hřebíčková & Graf, 2014; Jussim, Cain, Crawford, Harber & Cohen, 2009; Terracciano et al., 2005). The current study attempts to address the mixed nature of the existing literature concerning the racial/ethnic stereotype accuracy by combining it with the field of personality perception to see if actually being correctly identified as a member of a particular racial/ethnic group would result in being more accurately judged than if they had not been. This would suggest that the stereotype of the particular racial/ethnic group may possess some kernel of truth that is informative in the judgment of a particular individual.

### **Gender Stereotypes**

Comparatively speaking, the examination of gender stereotypes is a relatively new area of psychology. Gender stereotypes have only been of interest to stereotype researchers since the 1970s, with the addition of women in the field (Glick & Rudman, 2010). Like racial/ethnic stereotypes, gender stereotypes assume a set of characteristics are shared among members of a given gender, either in their entirety or at a greater likelihood than the general population. For example, women are assumed to show a greater amount of communal traits, such as nurturing and warmth, while men are assumed to show a greater number of agentic traits, such as activeness and competence (Swim & Hyers, 2009). Using a similar method to what was used to investigate the accuracy of these beliefs, individuals have been able to demonstrate at least some validity in their stereotypes (Jussim et al., 2009). When tested, people were generally accurate in their gender stereotypes in terms of personality, with the raters rating the typical man and typical woman being highly correlated with the observed means of men and women on specific traits (Löckenhoff et al., 2014). Similar results have been found for political and social beliefs as well, with people being able to estimate the percentage of men and women who would endorse particular beliefs on social and political issues quite well, particularly with women (Diekman et al., 2002). The current study builds upon this work by asking not whether people are able to accurately judge the traits and/or beliefs of the typical man or woman, but whether they would be able to accurately judge the traits of a particular individual, and whether knowing the gender of the individual influences the judgement's accuracy, as measured by agreement between the target and the rater.

### **Stereotype Use**

With a basic understanding of both gender and racial/ethnic based stereotypes, particularly in terms of their accuracy, one question that can be brought up is how such a phenomenon occurs. To put it another way, what is the mechanism by which the observed examples of stereotype accuracy found in the literature occur? Two possible avenues that this could work through are the aforementioned social projection and self-stereotyping. Social projection states that the individual uses some internal conception of a prototypical member of a given group or individual, often using the self as a form of reference (Cho & Knowles, 2013; Crawford et al., 2011). When dealing with intergroup interactions, the ideas of “prototypic individuals” could be treated as almost synonymous with the idea of stereotypes, with both being the standard with which an individual or group is compared (Hilton & von Hippel, 1996).

If one is to assume that the accuracy of the prototypes that one uses during social projection could differ in the real world, then the use of, or failure to use, these same conceptions could underlie the difference in social perception accuracy existent in the literature. According to this model, the observer would project a given stereotype on an unknown target based on their group membership or memberships while they are making a judgment about them (Cho & Knowles, 2013; Hilton & von Hippel, 1996). This stereotype would then serve as a baseline for the assessment of the traits of the person as a whole. If this stereotype reflects the actual traits of the target, as the “kernel of truth” hypothesis presupposes, then its use could help increase accuracy of the personality assessment. If social projection serves as a mechanism that results in stereotype accuracy, then the same traits that would impact an individual’s prejudice could also affect the accuracy of their social judgments. Consider the following scenario: a rater uses a

gender-based stereotype in their judgment of a target. This stereotype reflects the self-reported traits of the target relatively well. Those who are high in modern sexism and/or ambivalent sexism, or are more prejudiced in general maybe more likely to use gender-based stereotypes. If the stereotypes are accurate, then their use could cause individuals to make more accurate judgments. In the same way, if one were to be using an ethnic stereotype that matches the actual traits of an individual, then those more prone to express racial/ethnic prejudice, such as those higher in modern racism or colorblind racial ideology, or just more prone to prejudice in general could create more accurate judgments, as they should make use of the stereotype more often as well. However, if the information used in social projection was incorrect, use of stereotypes should lead to a decrease in the accuracy of perception.

Self-stereotyping, on the other hand, would place the source of the phenomenon of stereotype accuracy not necessarily in the accuracy of a stereotype being used by the observer, but in the targets themselves. Self-stereotyping states individuals who identify strongly as a member of a given social group will define themselves in terms of the prototypic member of the group (Cho & Knowles, 2013; Otten & Epstude, 2006). As mentioned previously, the conception that the individual has of the prototypic member of their own social group would match closely with any stereotype that may exist for the social group. This internalization of what is typical for a member of a given group could influence how the individual target see himself or herself on given traits, often aligned with those of the group (Otten & Epstude, 2006; Reid & Hogg, 2005). This explanation would imply that the accuracy of a judge's observations of an individual would be the result of the target identifying with the stereotype. The more that an individual target

identifies with belonging to particular social group, the more that he or she might behave in accordance with the stereotype of that group. Importantly, this would mean that the levels of prejudice in the rater should *not* be related to the stereotypicality, and hence accuracy, of the target's personality. While a direct comparison between the two mechanisms by which the stereotype accuracy literature may operation is beyond the scope of this study (see Cho & Knowles, 2013 for comparison), the current study predicts that participants will utilize social projection as a part of their judgments, projecting their concept of a prototypical member of the relevant social groups onto the target, which would in turn effect their perception of the individual.

### **Present Study**

This study set out to examine the influence of perceptive accuracy and expressive accuracy effects of stereotyping and prejudice on personality perception through short self-descriptions. To investigate effects of expressive accuracy, raters were given accurate or inaccurate gender labels, as well as accurate or inaccurate ethnic labels attached to targets' short written self-descriptions. To investigate the effects of perceptive accuracy, raters reported on their own levels of prejudice. Due to the complex nature of social perception, this study tested two different hypotheses, which requires the examination of six different variables of interest, consisting of two independent variables, three moderating variables, and one dependent variable. The independent variables consist of accuracy of gender labels of the target, which would theoretically run from completely inaccurate to completely accurate, and accuracy of racial and ethnic labels of the target, which would theoretically range from completely inaccurate to completely accurate. Manipulations of both of these independent variables would be predicted to

influence the dependent variable of accuracy of rater judgment. This would run from completely inaccurate to completely accurate. The relationship between the independent and dependent variables are predicted to be moderated by the variables of sexism, which varies from low to high sexism; level of racial prejudice, which ranges from low to high racial prejudice; and general level of prejudice, which ranges from low to high general prejudice.

This study examines two hypotheses. Hypothesis 1 states that there would be a positive relationship between having a correct or incorrect gender label and the accuracy of rater judgments of personality. Correct gender labels would result in more accurate judgments of personality, as measured by rater-target agreement on whole personality. This relationship was expected to be stronger among those who are higher in general prejudice and higher in sexism. Hypothesis 2 is similar to the first, that there would be a positive relationship between ethnic label accuracy and the accuracy of rater judgments of personality. This would be expected to be moderated by general prejudice and racism. Both hypotheses would suggest that the accuracy of these personality judgments occurs through social projection.

### **Preliminary Study**

In order to test the reasonableness of the hypotheses prior to running a full scale study, a preliminary study examined data previously published in Chan and Mendelsohn (2010). In the original study, undergraduate students from the University of California at Berkeley [UCB] (N=292) were brought into the lab to read and rate three unscripted dyadic Internet conversations selected from those generated in Williams and Mendelsohn (2008) as a part of a study on how gender influences text-based interaction and impressions. Williams and Mendelsohn (2008) had participants communicate over the Internet with an unknown partner and were instructed to “get to know your [the participant’s] partner” through a 15 minute communication.

The final sample of the data obtained from Chan and Mendelsohn (2010) was 61% female, with an ethnic composition of 44% Asian Americans, 27% European Americans, 7% Latino/a Americans, 2% African Americans, and 20 % Americans of other ethnic descent. Participants read the same three conversations, for a total of six targets with the individual targets either being labeled with both gender (“male, female”) and ethnic (African American”, “Asian American”, European American”) labels (e.g. “Person 1 is an African American female”), gender labels only, or no labels at all. After viewing each conversation, participants were asked to rate both targets in the conversation using a 31-item item personality measure drawn from the Adjective Check List (Gough & Heilbrun, 1983). They were selected to capture a range of gender-stereotypic and gender-neutral adjectives (Bem, 1974; Williams & Best, 1990), as well as a range of social desirability (Anderson, 1968; Gough & Heilbrun, 1983). A full alphabetical list of the adjectives used can be found in Appendix A. Raters were asked to



indicate their endorsement of each adjective on a 7-point Likert scale. After completing the ratings, participants were thanked and debriefed.

To determine the reasonability of the hypotheses of the current study, the targets that the participants had rated were coded in terms of the labels they were given and if they correctly identified the target (e.g. Was the Asian American female targets correctly labeled as a female? Asian American?). The labels were inaccurate gender label, accurate gender label, or missing gender label, as well as inaccurate ethnicity label, accurate ethnicity label, or missing ethnicity label; accurate label was operationalized to mean that the label matched the self-reported group membership of the target. The accuracy of personality ratings were assessed by correlating each observer's ratings of the target's whole personality profile with the self-reported personality of each of the six targets for a total of 1748 individual ratings, with each of the 292 participants rating six targets. The mean correlations between the participant ratings and the target self-ratings were then compared across the conditions. Post hoc Sidak-Bonferroni analyses reveal no statistical difference between the inaccurate ( $M=0.23$ ,  $SD=0.28$ ) and missing gender label conditions ( $M=0.23$ ,  $SD=0.25$ ) as well as between the inaccurate ( $M=0.26$ ,  $SD=0.26$ ) and missing ethnicity label conditions ( $M=0.23$ ,  $SD=0.26$ ). These analyses were conducted in order to minimize the incidents of Type I errors. If they were found to differ, we would be more certain that we found true effects. Therefore, the inaccurate and missing label conditions were collapsed for both gender and ethnicity for further analyses. We found a significant main effect of gender labels ( $F(1, 1721)=18.26$ , partial  $\eta^2 < .011$ ,  $p < .001$ ), such that those who received accurate gender labels demonstrated higher mean correlations between participant ratings and target self-ratings ( $M=.29$ ,  $SD=$

.27) than those who did not receive accurate gender labels ( $M=.23$ ,  $SD=.25$ ). No such effect was found for ethnicity labels ( $F(1, 1721) = .002$ , partial  $\eta^2 < .001$ ,  $p=.963$ ), with there being no mean difference between those who did receive accurate ( $M=.26$ ,  $SD=.26$ ) or did not receive accurate ethnicity labels ( $M=.26$ ,  $SD=.26$ ). In addition, no gender x ethnic label interaction ( $F(1, 1721) = .49$ , partial  $\eta^2 < .001$ ,  $p=.483$ ) was found. A subsequent independent sample T-test revealed that those in the accurate gender label condition were more accurate in their perceptions of the targets' personality profiles than those in the inaccurate gender label condition ( $t(1723) = 4.42$ ,  $p < .001$ ). These results provide evidence to suggest the feasibility of the current study, as it demonstrates participant ratings of text based communication can differ in accuracy depending upon whether the text was correctly labeled as belonging to a particular social group such as gender. It also demonstrates the feasibility of using simple labels as an experimental manipulation of assumed social group membership. The current study expands upon this by using a measure of personality based upon the Five Factor model, as well as being based upon text based communication created in a different context.

## Methods

### Participants and Sampling procedures:

The present study utilized a between-subjects experimental research design to test the relevant hypotheses previously discussed, that there would be a positive relationship between accuracy of gender and/or ethnic labels and the impressionistic accuracy of rater judgments of personality based on short self-descriptions, moderated by rater level of prejudice. Two samples were planned to be drawn, one being an undergraduate student sample from a public, mid-Atlantic university undergraduate population, as well as one being a crowdsourced community sample. For the undergraduate student sample, participants were recruited through the university subject pool. This population is selected for both recruitment convenience, as well as allowing a certain level of control that is required to run an experimental design easily. As this study examines the processes of making social judgments based on minimal information, there is no theoretical basis to suggest that an undergraduate population would differ from the general population in relevant characteristics that would influence the process. In addition, the shared social environment of a university would increase the likelihood that particular stereotypes for different groups are shared (Henry, 2008). Furthermore, the majority of studies in psychology have been performed using student samples since the 1960s, which would allow this study to better be compared with existing literature (Henry, 2008; Sears, 1986). Finally, the selected self-descriptions were produced by individuals from the same population, so members of the subject pool would have increased expertise in understanding the targets.

The population of the public mid-Atlantic university that this study uses differs from University of California at Berkeley in several important ways that makes it a very

fruitful population for which to conduct the study. The first major difference between the two institutions is size. In Fall 2015, the undergraduate enrollment of UCB was reported as 27,496 students (University of California Berkeley Office of Planning & Analysis [UCBOPA], n.d.). In comparison, the institution where the student sample will be drawn has 4,899 undergraduate students (Rutgers University Office of Institutional Research and Academic Planning [RUOIRAP], 2016), making it roughly one fifth of the size. This small population provides a more intimate environment for its students, allowing undergraduate students at the Mid-Atlantic institution to have more experience with interacting with a larger section of the total student body. This increased experience with other students may influence the ability of individuals to detect and utilize cues that may provide important information to be used in their perception of the targets (Funder, 1995, 2012; Letzring, 2015).

Another area that the populations differ is in the general racial ethnic breakdown of the institutions. UCB's undergraduate student body was approximately 39 % of Asian heritage, 3 % African American/Black, 13.7 % Latino/a and 26 % White in Fall 2015 (UCBOPA, n.d.), while the Mid-Atlantic institution's student body was 9.2 % Asian descent, 16 % African American, 12.8 % Latino/a, and 54.7 % White (RUOIRAP, 2016). Despite the smaller size, the Mid-Atlantic institution allows for a more racial/ethnic diverse sample, which may assist in demonstrate the generalizability of the results across racial/ethnic groups.

The crowdsourced community sample, which was to be used to replicate any findings discovered, was planned to be recruited from Amazon's Mechanical Turk (MTurk) users residing within the United States in exchange for \$1.00 in compensation.

MTurk, and other crowdsourcing websites like it, functions as an open online marketplace that allows registered researchers to post online experiments and other computer-based tasks and recruit participants to complete them (Gosling & Mason, 2015). MTurk especially boasts a sizable and diverse population of roughly 100,000 users from over 100 countries that complete thousands of tasks daily in exchange for minimal financial compensation, typically a few cents for a short survey (Buhrmester, Kwang & Gosling, 2011). This allows for researchers to be able to recruit large samples of participants within a matter of days with only minor financial investments (Buhrmester et al., 2011; Gosling & Mason, 2015).

Internet samples are generally found to differ from traditional samples in many characteristics that, although not necessarily representative of the general population, allow for greater generalizability to the research findings (Gosling & Mason, 2015; Gosling, Vazire, Srivastava, & John, 2004). For example, 71% of participants in traditional samples report being female, while 55 % participants in MTurk samples are female (Buhrmester et al., 2011). In addition, participants in MTurk samples have a statistically higher mean age ( $M = 32.8$  years,  $SD = 11.5$ ) than general internet samples ( $M = 24.3$  years,  $SD = 10.0$ ) (Buhrmester et al., 2011). Since one examination of internet samples found that 66% of US participants were not of traditional college age group (18-22 years) (Gosling, Sandy, John, & Potter, 2010). Reliance on stereotypes and prejudice has been found to be higher among older adults than younger adults, due to either cohort effects or decreased ability to inhibit unintentionally activated stereotypes (Radvansky, Copeland, & Von Hippel 2010). As a result, the MTurk sample may display a greater prejudice than the traditional sample, and thus be affected differently by the

manipulation. Together, the use of both a student sample as well as an MTurk sample would help to compensate for each other's limitations and increase external validity of the study.

## Measures

### Personality.

**Big Five Inventory:** Target personality on the factor level was measured by the Big Five Inventory (BFI), a 44-item self-report measure of personality designed to measure an individual's personality on the factors of the Five Factor Model of personality. These factors include Openness to Experience, which describes the complexity and variety of one's mental and experiential life; Conscientiousness, which describes one's level of social impulse control; Extraversion, which describes one's level of sociability and assertiveness; Agreeableness, which describes one's level of prosocial and communal inclinations; and Neuroticism, which describes one's predisposition to negative emotionality (John, Naumann, & Soto, 2008). The BFI utilizes a 5-point Likert Scale, ranging from *strongly disagree* (1) to *strongly agree* (5), where respondents are asked to express their level of agreement that given items accurately describes the respondent. For example, one item asks the respondent "I am someone who... Is helpful and unselfish with others" (See Appendix B). The BFI is designed to strike a balance between brevity and detail, and is one of the most widely used measures of personality using the Big Five Model (John et al., 2008). Its sub-scales measuring each of the factors have been found to have good reliabilities in United States and Canadian samples ( $\alpha$  ranges from 0.75 to 0.90; mean  $>$  0.80, John et al., 2008). The BFI also has shown convergent validity with the other measures of personality that use the Five Factor Model

(corrected mean  $r=.95$ ) (John et al., 2008). One limitation of this measure is also its greatest strength: its brevity. Being only 44 items, the BFI is unable to assess individuals' personality on the thirty facets, only on the five factors. Nevertheless, it is one of the widest used measures of personality in research (John et al., 2008).

*National Character Survey (NCS)*. Target personality on the facet level was measured by the National Character Survey, a 30-item self-report measure, originally developed for use in cross-cultural studies of personality stereotypes, consisting of a series of bipolar items with 5-point scales, with adjectives and/or phrases labeling each pole. Respondents are asked to indicate where along the continuum they picture the assigned target, in this case themselves. For example, one item asks the respondent where they fall between the extremes of "Somber, dull, sober" and "Happy, cheerful, joyous" (See Appendix C). Each item assesses one of the 30 facets present in the Five Factor Model of personality (Allik, Mõttus & Realo, 2010; Terracciano et al., 2005). This scale has been used successfully in samples from 49 separate countries. The individual facets displayed decent reliability across the global sample ( $\alpha$  ranges from 0.89 to 0.97; median = 0.94). Furthermore, the NCS was found to measure the Five Factors and their constituent facets as intended (Terracciano et al., 2005; Allik, Mõttus & Realo, 2010). The NCS items were found to mostly have sufficient congruence to its intended model (variable congruence coefficient ranges from coefficient = .70 to coefficient = .99; total congruence = .89).

In order to measure the impressionistic accuracy of personality judgments made by the raters, they rated the targets using the same NCS scale, with accuracy being determined via subject-rater agreement, a method previously used to determine the

accuracy of gender and age stereotypes (see Chan et al., 2012; Löckenhoff et al., 2014). Subject-rater agreement was determined by running an intraclass correlation (ICC) between the rater's responses to the 30 items of the NCS and the target's self-reported responses on the same scale. This allows for a direct comparison between the entire profiles provided by both the rater and the target.

### **Racism.**

Raters' self-reported levels of racism were assessed using three different scales: Modern Racism Scale (MRS) (McConahay, 1986), Color-blind Racial Attitude Scale (CoBRAS) (Swim, Aikin, Hall & Hunter, 1995), and The Intolerant Schema Measure-Racism (ISM-R) subscale (Aosved, Long, & Voller, 2009).

*Modern Racism Scale.* Modern racism is a form of racism that is understood to be a more complex and subtle form of anti-Black racism than traditional or “old fashioned” racism, recognizing that explicit prejudice against individuals primarily based on race or ethnicity is not socially acceptable. As a result, anti-Black racism has become more symbolic in nature, focusing on more socially acceptable means of expressing prejudice (Olson, 2009). Modern anti-Black racism was measured by the Modern Racism Scale (MRS) (McConahay, 1986), one of the most widely used measures of racism used in social science (Olson, 2009). The MRS is a 7-item self-report survey using a 5-point Likert Scale, ranging from *strongly disagree* (1) to *strongly agree* (5), where respondents were asked to expressing their level of agreement with a given item (Gamst, Liang & Der-Karabetian, 2011). One sample item reads “Discrimination against Blacks is no longer a problem in the United States” (See Appendix D) (McConahay, 1986). This scale has been shown to be quite reliable (Cronbach's  $\alpha = .82$ ) (Gamst et al.,



2011). It has also been found to be predictive of participants' willingness to hire a Black candidate when presented with an average résumé, with those who scored higher on the MRS being less likely to endorse hiring an alleged African American applicant than those who scored lower on the scale (McConahay, 1983). In addition, individuals' scores on this measure were positively correlated with opposition to busing children to desegregate Louisville, Kentucky, public schools during both a 1976 ( $r = .51$ ) and 1977 ( $r = .39$ ) survey (McConahay, 1986). Furthermore, MRS correlated positively with anti-Black sentiment measured by a Feeling Thermometer among both a community sample ( $r = .38$ ) as well as a longitudinal study involving university students ( $r = .44$ ) (McConahay, 1986). Some limitations of this measure are that it was designed to focus on anti-Black prejudice, which may not function in the same way as other forms of racial prejudice, and that this "Modern" Racism Scale was originally produced in the 1980s, during which time racism may not have taken the same form in contemporary society. Nevertheless, its role as one of the most used scales for a nonreactive measure of racism necessitates its use in order to best place this study in the context of the literature.

***Color-blind Racial Attitude Scale (CoBRAS)***. Color-blind racial attitudes are a form of racial prejudice that states that race should not, and does not, matter, and thus ignore or downplay the continued existence of racism in society. This was measured via a 20-item self-report Color-blind Racial Attitude Scale (CoBRAS) (Neville, Lilly, Duran, Lee, & Browne, 2000). CoBRAS is scored using a 6-point Likert scale. Respondents are asked to express their level of disagreement or agreement with each item ranging from *strongly disagree* (1) to *strongly agree* (6). Higher scores represents greater support for color-blind racial attitudes (Zou & Dickter, 2013; Neville et al., 2000). One sample item

is “White people are more to blame for racial discrimination than racial and ethnic minorities” (See Appendix E) (Neville et al., 2000). This scale has been shown to have good internal reliability ( $.84 < \text{Cronbach's } \alpha < .91$ ), as well as test-retest reliability ( $r=.68$ ) (Gamst et al., 2011). In addition, it has been found to be correlated with other measures of racial attitudes [e.g. MRS ( $.36 \leq r \leq .55$ )], but not with social desirability, as measured by Marlowe-Crowne Social Desirability Scale ( $r = .13$ ) (Neville et al., 2000). This measure is especially useful in that it examines racial prejudice irrespective of any individual race and/or ethnicity.

***Intolerant Schema Measure (ISM)-Racism (ISM-R) subscale.*** The Intolerant Schema Measure (ISM) is designed to measure multiple forms of intolerant beliefs simultaneously in a single measure (Aosved et al., 2009). It is a 54-item self-report measure consisting of six subscales of 9-items each, derived from items from existing measures. To measure general racism, the 9-item Intolerant Schema Measure-Racism (ISM-R) subscale was used, with each item scored using a 5-point Likert scale, ranging from *strongly disagree* (1) to *strongly agree* (5). One sample item from the subscale is “If a racial minority family with about the same income and education as I have moved in next door, I would mind a great deal.” (See Appendix F) (Aosved et al., 2009). The subscale has been shown to have good internal reliability (Cronbach's  $\alpha = .82$ ). It was also shown to be significantly correlated with both the Old Fashioned Racism Scale ( $r = .84, p = .0001$ ) and ( $r = .90, p = .0001$ ), as well as Social Dominance Orientation ( $r = .65, p = .001$ ) (Aosved et al., 2009). This measure is relatively new, but its ability to measure racism without reference to individual races and/or ethnicities makes its use important when dealing with prejudice involving multiple racial/ethnic groups.

## **Sexism.**

Raters' self-reported levels of sexism were assessed using four different scales: Ambivalent Sexism Inventory (Glick & Fiske, 1996), Ambivalence Towards Men Inventory (Glick & Fiske, 1999), Modern Sexism Scale (Swim, Aikin, Hall & Hunter, 1995), and The Intolerant Schema Measure (ISM)-Sexism subscale (Aosved et al., 2009).

***Ambivalent Sexism Inventory.*** Ambivalent sexism is a form of prejudice based upon sex that examines the complexity of sexism, in terms of its positive (“Benevolent”) and negative (“Hostile”) components. Ambivalent sexism was assessed by the 22-item self-report Ambivalent Sexism Inventory (ASI) (Glick & Fiske, 1996). The ASI is scored using a 6-point Likert scale, with each point labeled, which respondents are asked to express their level of disagreement or agreement with each item. It ranges from *disagree strongly* (0), *disagree somewhat* (1), *disagree slightly* (2), *agree slightly* (3), *agree somewhat* (4), and *agree strongly* (5). One sample item reads “No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman” (See Appendix G) (Glick & Fiske, 1996). This measure has been shown to be sufficiently reliable across samples ( $.83 < \text{Cronbach's } \alpha < .92$ ). ASI scores were also found to positively correlate with other measures of sexism, such as Attitudes Toward Women Scale ( $r=.63, p<.01$ ), Old-Fashioned Sexism ( $r=.42, p<.01$ ), Modern Sexism ( $r=.57, p<.01$ ), and Rape Myth Acceptance Scale ( $r=.54, p<.01$ ). The ASI also was found to be predictive of individuals' endorsement of ambivalent attitudes towards women and adoption of stereotypes (Glick & Fiske, 1996).

***Ambivalence Towards Men Inventory.*** The Ambivalence Towards Men Inventory (AMI) is the sister scale to the ASI, examining both benevolent and hostile

gender-based prejudice against men, using a 20-item measure. The AMI is scored using a 6-point Likert scale, with each point labeled. Respondents are asked to express their level of disagreement or agreement with each item. It ranges from *disagree strongly* (0), *disagree somewhat* (1), *disagree slightly* (2), *agree slightly* (3), *agree somewhat* (4), to *agree strongly* (5). An example item from the inventory is “Men are less likely to fall apart in emergencies than women are.” (See Appendix H) (Glick & Fiske, 1999). The scale has been shown to have good internal consistency reliability ( $.83 < \text{Cronbach's } \alpha < .87$ ). It has also been shown to correlate positively with other measures of attitudes towards men, as well as with ASI (female respondents  $r = .76, p < .01$ ; male respondents  $r = .69, p < .01$ ). This measure serves as a counterbalance towards the ASI, in order to capture both prejudice against men and women.

***Modern Sexism Scale.*** Modern Sexism is a form of sexism that is characterized by the denial that discrimination based upon sex still occurs and antagonism towards any policy or movement to combat said discrimination, similar to modern racism’s denial of the continued existence of racism. As the concepts of modern sexism and modern racism are similar, The Modern Sexism Scale (Swim et al., 1995), which measures gender-based prejudice, is based on the Modern Racism Scale (McConahay, 1986). It consists of an 8-item self-report measure scored using a 5-point Likert scale, ranging from *strongly agree* (1) to *strongly disagree* (5). The scale consists of such items as “Discrimination against women is no longer a problem in the United States” (See Appendix I) (Swim et al., 1995). The scale was shown to be internally reliable ( $\alpha = .84$ ). In addition, the Modern Sexism Scale was found to be significantly correlated with Humanitarian-Egalitarian Values (female:  $r = -.29, p \leq .05$ ; male:  $r = -.16, p \leq .05$ ) (Swim et al., 1995).

***Intolerant Schema Measure (ISM)-Sexism.*** The Intolerant Schema Measure-Sexism (ISM-S) subscale is a 9-item subscale of the same measure that ISM-R was taken from. While ISM-R is a measure of racism, ISM-S is a measure of sexism. As with the ISM-R, each item was scored using a 5-point Likert scale, ranging from *strongly disagree* (1) to *strongly agree* (5), in order to measure sexism (Aosved et al., 2009). Individuals rate their level of agreement with the individual items, such as “There are many jobs in which men should be given preference over women in being hired or promoted” (See Appendix J) (Aosved et al., 2009). This scale shows good reliability ( $.82 \leq \text{Cronbach's } \alpha \leq .84$ ). Its content validity is demonstrated by its significant correlation with preexisting measures of sexism such as the Neosexism scale (Tougas, Brown, Beaton, & Joly, 1995) ( $r = .83, p = .0001$ ) and the Attitudes Toward Women Scale (Spence, Helmreich, & Stapp, 1973) ( $r = .83, p = .0001$ ), as well as being significantly correlated with Social Dominance Orientation ( $r = .48, p = .0001$ ) (Aosved et al., 2009).

### **General Prejudice.**

***Social Dominance Orientation (SDO) Scale.*** Social Dominance Orientation (SDO) (Pratto, Sidanius, Stallworth, & Malle, 1994) is an individual difference that describes an individual's support for maintaining social inequality among social groups, with those high in it endorsing the use of different legitimizing beliefs and myths to support these between group inequalities (Olson, 2009; Pratto et al., 1994). While technically not measuring prejudice, SDO is still one of the most common and prototypical measures of general prejudice, one by which other prejudice measures are often compared (Olson, 2009; Schneider, 2004). Social Dominance Orientation is measured by the Social Dominance Orientation Scale, a 16-item self-report measure.

Respondents are asked to rate their positive and/or negative feeling towards a series of statements on a 7-point Likert scale, with each point labeled. It ranges from *very negative* (1), *negative* (2), *slightly negative* (3), *neither positive nor negative* (4), *slightly positive* (5), *positive* (6), to *very positive* (7). This scale included such items as “To get ahead in life, it is sometimes necessary to step on other groups.” (See Appendix K) (Pratto et al., 1994). The scale has shown to be internally reliable for both the 16-item ( $\alpha=.91$ ) and 14-item forms (average  $\alpha= .83$  across samples), and demonstrated significant test-retest reliability over a 3-month interval ( $r=.81$ ,  $p < .01$ ). In addition, it has been found to be negatively correlated with a measure of tolerance (ranging from  $r= -.27$  to  $r=-.36$ , all  $ps < .01$ ). It has also been found to correlate with political-economic conservatism, a related concept, (average  $r =.38$ ), as well as predict attitudes towards political policies, even after taking conservatism into account (39 out of 41 items maintaining significance beyond at least  $p < .05$  after controlling for political conservatism). Furthermore, it strongly correlated with several measures of anti-Black prejudice (ranging from  $r= .42$  to  $r=.65$  and averaging  $r=.55$ ) as well as anti-Arab racism ( $r=.22$ ,  $p < .05$ ), and sexism (average  $r= .47$ ) (Pratto et al., 1994).

## **Procedure**

### **Undergraduate Sample.**

Undergraduate student participants (N=104, 76 [73.1%] females, 25 [24%] males, and 3 [2.9%] declined to state), took part in the study in exchange for partial course credit. The ethnic breakdown of participants was as follows: 19 (18.3 %) African American/Black, 13 (12.5%) Asian Americans, 47 (45.2%) European American/Whites, 8 (7.7%) Latin Americans, 9 (8.7%) Multi-Racial individuals, 1(1%) Native

American/American Indian, 1 (1%) Native Hawaiian American, 4 (3.8%) Others, and 2 (1.9%) failed to disclose. Participants were brought into the lab and given informed consent by a research assistant, being told that they would be taking part in a study of the perception of strangers' personality based on a short self-reported memory about a time that they felt happy. The selection of a self-reported memory of happiness to use as the target was based upon research that suggests that individuals are less accurate in their thin-slice judgments of others when experiencing sadness or depression (Ambady & Gray 2002). By selecting a happy memory, we hoped to limit the amount of sadness that the participants may have experienced as a part of this study, and thus maximize their potential for making accurate judgments.

During the informed consent procedures, participants were reminded that their participation was entirely voluntary and they would be allowed to terminate their session at any point, and their data not used, without penalty. In addition, some basic information was asked of them, in order to measure rater differences in personality perception. Upon giving informed consent, participants were taken into another room and sat down at a computer preset with the study in an online format, which was run through Qualtrics, an online survey service. The use of an online data collection service allowed for the more effective collection of data.

Once beginning the study, the participants were shown short descriptions about a time that an unknown individual felt happy (Mean word count=36). These self-descriptions were produced by participants from a prior study (see Appendix L). Each one was labeled as having been ostensibly written by a member of one of four racial/ethnic groups ("African American", "Asian American", "European American", and

“Latin American”) and one of two gender groups (“female”, “male”), for a total of eight possible label combinations. Each label combination matched the self-reported demographics of one of the targets. Each participant was shown the same eight short self-descriptions, but the order of labels of the individual ones differed for each participant and counterbalanced. However, each label was shown to every participant exactly once during their session, so that every label and self-description was used. The participants were shown two self-descriptions with both accurate gender labels and ethnic labels; two self-descriptions with accurate gender labels and inaccurate ethnic labels; two self-descriptions with inaccurate gender and accurate ethnic labels; and two self-descriptions with inaccurate gender labels and ethnic labels. After each self-description, the participants were asked to rate the target using the National Character Survey (NCS). They repeated this process for the other seven self-descriptions. After filling out the final individual’s NCS rating, the participant was asked to fill out a series of prejudice measures: Modern Racism Scale (MRS) (McConahay, 1986), Color-blind Racial Attitude Scale (CoBRAS) (Swim, Aikin, Hall & Hunter, 1995), The Intolerant Schema Measure (ISM)-Racism subscale (Aosved et al., 2009), Ambivalent Sexism Inventory (ASI)(Glick & Fiske, 1996), Ambivalence Towards Men Inventory (AMI) (Glick & Fiske, 1999), Modern Sexism Scale (MSS) (Swim, Aikin, Hall & Hunter, 1995), ISM-Sexism subscale (Aosved et al., 2009), and Social Dominance Orientation Scale (SDO). In addition, the participants were asked to fill out the BFI and the NCS to describe themselves, as well complete some sociodemographic questions. Upon finishing, the participants were thanked their help in the conduction of the study and debriefed about its nature.



Participants were given the chance to ask any question that they might have about the study, as well as a final chance for them to withdraw.

*Crowdsourced Sample.*

The crowdsourced sample was to follow a similar procedure to the undergraduate student sample. Participants would be recruited for the study through a posting on Amazon Mechanical Turk in exchange for a small monetary compensation of \$1.00. Participants would be shown a screen that contains the information in the informed consent sheet provided to the participants from the traditional sample, and would be asked to read the form prior to beginning the study. They would have been told that they would be taking part in a study on the perception of a stranger's personality based on a short self-reported memory about a time that said stranger felt happy. They would have been reminded that their participation is entirely voluntary and that they would be allowed to terminate their session at any point, and their data not used, without penalty. In addition, some basic information would be asked of them, in order to measure rater differences in personality perception. After the participants indicate that they have read and understood the informed consent information, they would be allowed to continue onto the study itself, housed through Qualtrics, allowing for ease of data collection and creating standardized experiences between the samples.

Upon beginning the study, the participants would have undergone the same procedure as the student sample. They would be shown a short description about a time that an unknown individual felt happy (Mean word count=36), that was produced by a participant from a prior study. Each one would be labeled as having been ostensibly written by a member of one of the four racial/ethnic groups ("African American", "Asian

American”, “European American”, and “Latin American”) and one of the two gender groups (“female”, “male”), for a total of eight possible labels. Each participant would be shown the same eight short self-descriptions, but the order of labels of the individual ones would differ for each participant, counterbalanced. They would be shown two self-descriptions with both accurate gender labels and ethnic labels; two self-descriptions with accurate gender labels and inaccurate ethnic labels; two self-descriptions with inaccurate gender and accurate ethnic labels; and two self-descriptions with inaccurate gender labels and ethnic labels. Each label would be shown to every participant exactly once during their session, so that every label and self-description would be used. After each self-description, the participants would be asked to rate the target using the NCS. They would repeat this process seven more times for the other seven self-descriptions. After filling out the final target’s NCS rating, the participants would be asked to fill out a series of prejudice measures: MRS, CoBRAS, ISM-Racism subscale, ASI, AMI, MSS, ISM-Sexism subscale, and SDO. In addition, the participants would be asked to fill out the NCS to describe themselves, as well as some sociodemographic questions. Upon finishing, the participants would be thanked for their help in the conduction of the study and debriefed about its nature, and given contact information of the Principal Investigators in order to contact them if they have any questions about the study. The participants would also be given the code required for them to claim their compensation. The crowdsourced sample, however, was not collected, due to several factors that will be discussed later in the Results section.

## Results

### Scale Development and Descriptive Statistics

**Gender label accuracy.** In order to test the hypotheses that there would be a positive relationship between the accuracy of gender labels and the impressionistic accuracy of rater judgments of personality based on short self-descriptions, each self-description was coded as having accurate or inaccurate gender information. This was achieved by comparing the gender label that the participant was given for each self-description with the self-reported gender of the self-description's author. The label was coded as "Accurate" when the label and the gender of the self-description's author matched. The label was coded as "Inaccurate" when the label and the gender of the self-description's author did not match.

**Racial/ethnic label accuracy.** In order to test the hypotheses that there would be a positive relationship between the accuracy of racial/ethnic group labels and the impressionistic accuracy of rater judgments of personality based on short self-descriptions, each self-description was coded as having accurate or inaccurate racial/ethnic group information. This was achieved by comparing the racial/ethnic group label that the participant was given for each self-description with the self-reported racial/ethnic group membership of the self-description's author. The label was coded as "Accurate" when the label and the racial/ethnic group of the self-description's author matched. The label was coded as "Inaccurate" when the label and the racial/ethnic group of the self-description's author did not match.

**Impressionistic accuracy.** To calculate the dependent measure of impressionistic accuracy, an ICC was ran between the participants' rating profiles of each of the self-

descriptions and the self-reported profile provided by the targets after they wrote the self-description. (mean  $r = 0.13$ ,  $SD = .32$ ). Of the self-descriptions, Self-Description E displayed the highest mean level of rater-target agreement ( $M = .40$ ,  $SD = .23$ ), while the ratings of Self-Description G displayed the lowest mean level of rater-target agreement ( $M = -.26$ ,  $SD = .22$ ). Besides Self-Description G, only Self-Description B displayed a negative mean level rater-target agreement ( $M = -.04$ ,  $SD = .27$ ). All other Self-Descriptions had positive mean rater-target agreement (Table 1).

**Prejudice.** To create indices for sexism, racism, and general prejudice, we ran a principal components analysis (PCA) on the measures used in the study. For the racism measures, the PCA revealed a single factor above an Eigenvalue of 1.00, accounting for 71.26% of the variance. The PCA that was ran for the sexism measures revealed two factors above an Eigenvalue of 1.00, with the first factor accounting for 52.58% of the variance, and the second accounting for 29.02 % of the variance. To examine how much the different prejudice measures covary, a correlation matrix was run for all of the prejudice measures to assess degree of relationship among them. Due to the high correlation among the measures (Table 2), an additional PCA was run to analyze all of the prejudice measures as a whole. The PCA revealed two factors above an Eigenvalue of 1.00, with the first factor accounting for 47.78% of the variance and the second factor accounting for 21.00 % of the variance. Only AMI (Factor Loading=.90) and ASI (Factor Loading=.76) loaded onto the second factor to any significant degree.

To create the indices we standardized the responses on the individual prejudice measures by converting them into z-scores. We then calculated the arithmetic mean of the individual standardized measures of each type to produce the three aforementioned

measures of racism, sexism, and global prejudice. The Racism index was formed from mean of standardized transformations of MRS, CoBras, and ISM-R. The Sexism index was formed from the mean of standardized transformations of ISM-S, MSS, AMI, and ASI. Global Prejudice was obtained by the mean of standardized transformations of SDO, ISM-S, MSS, AMI, ASI, ISM-R, CoBRAS, and MRS. This allowed for the controlling for prejudice in correlations between the accuracy of the individual gender and racial/ethnic group information provided and the accuracy of participant ratings. This will be revisited during our discussion of the influence of prejudice.

### **Effect of Gender Labels**

Hypothesis 1 predicts a significant positive relationship existing between accuracy of gender labels and accuracy of an individual's perception of an unknown author of a short self-description. To examine this hypothesis, a 2x2 Gender Label x Ethnic Label analyses of variances (ANOVAs) were conducted, with the rater-target agreement for self-descriptions separately serving as the dependent variable. The analysis found no significant main effect of gender labels ( $F_s(1,100) = .01$  to  $2.29$ , partial  $\eta^2 = .000$  to  $.022$ , all  $ps > .133$ ) for any of the self-descriptions. Interaction effects will be discussed in the following section on the effect of racial/ethnic labels.

### **Effect of Racial/Ethnic Labels**

Hypothesis 2 predicts a significant positive relationship existing between accuracy of gender labels and accuracy of an individual's perception of an unknown author of a short self-description. This hypothesis was tested using the same 2x2 Gender Label x Racial/Ethnic Label ANOVAs to test Hypothesis 1. This analysis found a marginally significant main effect of Racial/Ethnic label for Self-Description H ( $F$

(1,100) = 3.89,  $\eta^2 = .037$ ,  $p = .051$ ), which is depicted in Figure 1. A post hoc independent samples t-test on Self-Description H was performed, revealing a marginally significant relationship in the expected direction ( $t(102) = 1.91$ ,  $p = .059$ ), with participants who were provided with an accurate Racial/Ethnic Label for Self-Description H displaying a slightly greater mean rater-target agreement ( $M = .23$ ,  $SD = .25$ ) than those provided with an inaccurate Ethnicity Label ( $M = .14$ ,  $SD = .22$ ). However, all other stimuli failed to show a main effect of Racial/Ethnic label of similar significance ( $F_s(1,100) = .11$  to  $2.33$ , partial  $\eta^2_s = .001$  to  $.023$ , all  $p_s \geq .129$ ).

### **Interaction Effects**

An examination of the interaction between Gender Label and Racial/Ethnic Label was conducted. Figure 2 shows a significant Gender Label by Racial/Ethnicity Label interaction between was found for Self-Description A ( $F(1,100) = 4.08$ , partial  $\eta^2 = .039$ ,  $p = .046$ ). An inspection of the means for Self-Description A found participants who were shown both inaccurate Gender Labels and Racial/Ethnic Labels displayed the greatest mean level of rater-target agreement ( $M = .20$ ,  $SD = .28$ ), with those being shown both accurate Gender Labels and Ethnicity Labels displaying slightly lower mean levels of rater-target agreement ( $M = .17$ ,  $SD = .25$ ). In addition, a marginally significant Gender Label by Racial/Ethnic Label interaction was found for Self-Description C ( $F(1,100) = 3.75$ , partial  $\eta^2 = .036$ ,  $p = .056$ ), which is shown in Figure 3. Participants who were given inaccurate Gender Labels and accurate Racial/Ethnic Labels displayed the greatest mean level of rater-target agreement ( $M = .35$ ,  $SD = .22$ ), with those given an accurate Gender Label and inaccurate Ethnicity Label displaying slightly lower mean levels ( $M = .31$ ,

SD=.31) (Table 3). All other relationships were nonsignificant ( $F$ s (1,100) = .11 to 1.94, partial  $\eta^2$ s = .001 to .026,  $p$ s  $\geq$  .167).

### **Collapsed Effect of Labels**

**Main effects.** With the inconsistency of the previous results across the individual self-descriptions, we decided to collapse the results across self-description and treat the individual self-descriptions as a third independent variable. We then ran a 2 x 2 x 8 Gender Label x Ethnicity Label x Self-Description ANOVA, the results of which are shown in Figure 4. We found a significant main effect of Self-Description ( $F(7,800)$  = 69.51, partial  $\eta^2$  = .378,  $p$  < .001). This new ANOVA also found a marginally significant effect for Racial/Ethnic Labels ( $F(1,800)$  = 3.66, partial  $\eta^2$  = .005,  $p$  = .056). A post hoc independent t-test found the relationship to be in the expected direction. ( $t(830)$  = 2.00,  $p$  = .045), with those provided with an accurate Racial/Ethnic Label displaying greater mean rater-target agreement ( $M$  = .16,  $SD$  = .32) than those provided with an inaccurate Racial/Ethnic Label ( $M$  = .11,  $SD$  = .32). However, we did not find a significant main effect of Gender Label ( $F(1,800)$  = .001, partial  $\eta^2$  < .001,  $p$  = .970).

**Interaction effects.** In addition to the above main effects, we found a significant three-way interaction among Gender Labels, Racial/Ethnic Labels, and Self-Description ( $F(7,800)$  = 2.19, partial  $\eta^2$  = .019,  $p$  = .033). This interaction is shown in Figure 4. As a result, it is the combination of these variables that change the relationships (Table 4). No other significant interaction effects were found (Gender Labels x Ethnicity Labels ( $F(1,800)$  = .25, partial  $\eta^2$  < .001,  $p$  = .620), all other  $F$ s(7,800) = .99 to 1.44, partial  $\eta^2$  s = .009 to .012,  $p$ s  $\geq$  .186).

### **Prejudice**

Further analyses of the results attempted to control for levels of sexism, racism, and general prejudice in the relationships between the accuracy of labels and rater-target agreement. For the uncollapsed data, racism was only significantly correlated with rater-target agreement ratings of Self-Description B ( $r = -.24, p = .015$ ) and Self-Description E ( $r = -.27, p = .006$ ), but in the nonpredicted direction. Like Racism, Global Prejudice significantly correlated with self-other agreement ratings of Self-Description B ( $r = -.21, p = .032$ ) and Self-Description E ( $r = -.25, p = .011$ ), but in the nonpredicted direction. No significant relationships were found for sexism (all  $r$ s =  $-.14$  to  $-.001, p$ s  $\geq .170$ ). Upon collapsing the data across Self-Descriptions, we ran another correlation between the standardized prejudice measures and general accuracy levels, finding a significant negative relationship, as shown in Figure 5 between the two variables for Global Prejudice ( $r = -.09, p = .013$ ). We found marginally significant negative relationships with general prejudice for both Sexism ( $r = -.07, p = .057$ ) and Racism ( $r = -.07, p = .057$ ).

### **Post hoc Analyses**

**Factor accuracy.** To examine if raters differed in their ability to accurately perceive individual personality factors, a post hoc series of 2 x 2 x 8 Gender Label x Ethnicity Labels x Self-Description ANOVAs were conducted, one for each individual personality factor. A significant main effect of Self-Description was found for each of the factors ( $F(7,800)$ s = 5.50 to 25.86, partial  $\eta^2$ s = .046 to .185,  $p$ s < .001). Significant two-way interactions between Gender Label Accuracy and Racial Ethnic Label ( $F(1,800) = 3.91$ , partial  $\eta^2 = .005, p = .048$ ), as well between Gender Label Accuracy and Self-Description ( $F(7, 800) = 2.76$ , partial  $\eta^2 = .024, p = .008$ ) were found for Conscientiousness were found. A significant three-way interaction among Gender Label,



Ethnicity Label, Self-Description was found for Openness to Experience ( $F= 3.91$ , partial  $\eta^2 = .033$ ,  $p < .001$ ). No other effects were found to be significant (See Table 5-Table 8). Prejudice was not found to significantly correlate with measures of accuracy on any of the factors ( $r_s = -.02$  to  $.02$ ,  $p_s \geq .234$ ).

**Rater personality.** Due to previous research suggesting that observer personality may influence accuracy of social judgments (Letzring, 2008; Letzring, 2015), we decided to examine if rater personality influenced the impressionistic accuracy of judgments for the current study. Analysis using the BFI was not performed due to a mechanical error resulting in participants not seeing all of the items. As a result, participants' self-reported responses on the NCS were used as a measure of personality. Self-reported levels of rater Neuroticism was significantly correlated negatively with rater-target for Self-Descriptions B ( $r = -.23$ ,  $p = .022$ ), Self-Description C ( $r = -.23$ ,  $p = .022$ ), and Self-Description E ( $r = -.23$ ,  $p = .017$ ). Self-reported levels of Extraversion was positively associated with rater-target of judgments of Self-Description E ( $r = .21$ ,  $p = .036$ ), but was negatively associated with judgments of Self-Description G ( $r = -.29$ ,  $p = .003$ ). Agreeableness was positively associated with rater-target for Self-Description B ( $r = .24$ ,  $p = .014$ ), Self-Description C ( $r = .38$ ,  $p < .001$ ), Self-Description D ( $r = .28$ ,  $p = .004$ ), but negatively associated with Self-Description F ( $r = -.26$ ,  $p = .007$ ) and Self-Description G ( $r = -.33$ ,  $p = .001$ ). Conscientiousness is positively associated with rater-target agreement for Self-Description C ( $r = .32$ ,  $p = .001$ ), but negatively associated with Self-Description G ( $r = -.27$ ,  $p = .006$ ) (See Table 10 for Correlation Matrix).

When controlling for Global Prejudice, we found a similar pattern of relationships among rater personality and self-other agreement among the different self-descriptions.

However, in addition to rater Conscientiousness being positively significantly associated with rater-target agreement of Self-Description B ( $r=.20, p=.039$ ) and negatively associated with Self-Description G ( $r= -.29, p= .003$ ). It was also positively associated with rater-target agreement for Self-Description C ( $r=.36, p < .001$ ), Self-Description D ( $r=.20, p= .042$ ), and Self-Description E ( $r= .24, p= .015$ ) (See Table 11 for complete Correlation Matrix).

When collapsing across self-descriptions, we found rater Neuroticism was significantly correlated negatively with accuracy ( $r= -.08, p = .017$ ). We also found raters' self-reported levels of Agreeableness ( $r= .09, p =.012$ ) and Conscientiousness ( $r= .08, p =.016$ ) were significantly associated positively with accuracy, but not rater Extraversion ( $r= .04, p =.238$ ) or Openness to Experience ( $r= .01, p =.877$ ) (See Table 12 for correlation matrix). These relationships follow a similar pattern when controlling for prejudice (See Table 13 for correlation matrix).

### **Crowdsourced Sample**

The collection of the planned crowdsourced sample, which was to be drawn from MTurk, was dropped due to multiple factors. The first factor that led to the decision to drop the planned data collection was one of purpose. The purpose of the MTurk sample was to be collected as a replication of the traditional sample and any of its findings. The traditional sample, however, failed to demonstrate consistent findings. Among the analyses of the effects of only Gender Labels and Racial/Ethnic Labels, only one significant finding can be found, which was an interaction between the two variables for Self-Description A ( $F(1,100) = 4.08, \text{partial } \eta^2=.039, p = .046$ ). Even if marginally significant results are considered, only three more effects can be discovered, with a mean of partial  $\eta^2=.029$ . This by itself is not too much of an issue, as to capture an effect of this

magnitude with sufficient power ( $1-\beta=.80$ ) one would have to have collect a sample of  $N=816$ . However, to properly detect the smallest effect of interest found (partial  $\eta^2=.005$ ), the sample would have to be quite large ( $N= 2872$ ) due to the weakness of the effect (Lakens, 2013). A sample of this size would be difficult to obtain in sufficient time for analysis.

This issue of time is the second factor that led to the decision to drop the planned collection of MTurk data. The current study was performed as part of a Master's research project; as a result, the study was limited in the time available for data to be collected and analyzed. Due to the inexperience of the author, the process of data cleaning and analysis took longer than initially expected, leaving the time available left for MTurk data to be collected and cleaned enough to be usable for comparison limited. In addition, the previously discussed inconsistency of the findings hints at the possible need to revisit the procedure in order to account for additional confounds. In order to not compromise the quality of the research due to the condensed time frame and the necessary large size of the sample size, it was decided to drop the MTurk data collection, with the possibility to revisit it at a later date once these limitations could be addressed.

## Discussion

This present study set out to examine the role that being given an accurate gender and/or ethnicity label have on their perceptions of an individual, based solely on a short self-descriptions. We hypothesized that being given an accurate label of either type would result in an increased level of accuracy. In addition, we hypothesized that levels of sexism, racism and global prejudice would moderate the relationship. However, the current study failed to find support for the hypotheses. On both the general and stimulus specific levels, no significant main effects were found for accuracy of Gender Labels or the accuracy of Racial/Ethnic Labels.

However, the presence of a marginally significant relationship between Racial/Ethnic Label accuracy and general level of rater-target agreement, as well as a similar relationship found for a single self-description, does seem to lend at least some support for the existence of a possible relationship, albeit very weak. For example, the effect size of ethnic label accuracy for general self-other agreement is partial  $\eta^2 = .005$ , which is relatively small. The small effect size, as discussed in Results, makes the rejection of a null hypothesis difficult in except in larger samples, as it would only have a small influence on the participants that may be mistaken for error without enough participants. This is for the effects of this relationship may hold very little practical significance by itself. This provides support for the idea that there are likely other more influential factors that are at play in making when making an accurate social perception of a stranger based in a short self-description than the stranger's perceived racial/ethnic group.

The influence of other confounding factors besides perceived social group membership is best demonstrated by the fact that the specific target that is being rated is the only significant main effect for general self-other agreement. While the purpose of this study was to assess the perceptive accuracy component of impressionistic accuracy, which is to say that we adjusted how judges are able to perceive different targets, the differences among the different targets in terms of how accurately they were perceived is the result of the expressive accuracy of the original targets themselves (Biesanz, 2010). Each of these self-descriptions may differ in the various cues about the personality of its author that they provide to the rater. As a result, some may have provided more useful cues to the rater than others. The three-way interaction among gender labels, ethnicity labels, and self-description could be explained in this way. It is in context of all three available elements that the rater made his or her assessment.

A possible limitation of the study deals with how language is naturally used. During the planning of the study, the eight stimuli were selected due to their lack of any direct references to the gender or racial/ethnic group membership or seem stereotypic of any one group. This was done so as to make the self-descriptions appear to have been plausibly written by any individual, and thus make the manipulation effective. However, it is possible that there were some subtle cues in their structure or word choice that may have betrayed more about the group membership of the targets than anticipated. Previous research has found that there are differences in word use and other linguistic behaviors between genders and across age groups that are able to be detected by a sophisticated program, such as Linguistic Inquiry and Word Count (LIWC) (Newman, Groom, Handelman, & Pennebaker 2008; Pennebaker, Francis, & Booth, 2001; Pennebaker et

al., 2003). It is plausible that participants picked up on these unseen cues, which weakened the effect of the manipulation. However, a study on online communication that had some participants actively pretend to be a member of the other sex during the online interaction, participants were limited in their ability to avoid being deceived in the gender of their unseen partner, which suggests that their ability to utilize these unseen cues are limited as well (Williams & Mendelsohn, 2008).

Due to the consistently significant effects of Self-Descriptions on the level of rater-target agreement in the study, a cursory inspection of the individual self-descriptions was done in order to speculate on any possible areas of interest that may describe the results. When the three self-descriptions with the greatest mean rater-target agreement was investigated, it would appear that two may have commonality. Both Self-Description E ( $M=.40$ ,  $SD=.23$ ) and Self-Description C ( $M=.27$ ,  $SD=.28$ ) describe a scenario that is relatively intimate, being arguably the most emotionally intimate of the stimuli; while most of the other self-descriptions discuss such topics as winning a soccer game or making jokes, Self-Description C's author discusses watching his or her child being inducted into an honor society and the pride that they felt and Self-Description E discusses graduating from community college and the pride and excitement that the author felt, despite his or her family being unable to make it. Both of these stimuli talk about a great deal of the authors' internal states, which may have impacted their level of agreement positively by increasing the information available to the raters. Future research may need to take into account the variability in the level of self-disclosure in the selection of stimuli, so as to balance the need to balance the needs of experimental control and maintaining ecological validity.

A second intriguing similarity that can be speculated upon is in the role of children in both Self-Description C and Self-Description E and what that may have interacted with the sample. Of the eight self-descriptions, only these two directly reference the author possibly having children, and are among the most explicit in reference to family as a whole. The presence of close family in these self-descriptions may have primed a sociability focus in participants. While still not unanimous, personality perception literature tends to suggest that female observers are slightly more accurate in their perceptions of strangers than male observers, especially among those who score high on sociability and are interpersonally warm (Letzring, 2008, Letzring, 2010; Li & Chignell, 2010). As a result, the primarily female (73.1% of female) sample may have been primed to be more socially focused while rating these self-descriptions, and thus more prone to be accurate in their assessments.

With the large amount of female participants in the study, one could argue that the lack of consistent findings as a part of the study would suggest the nonexistence of an actual effect. If female participants are deemed as more accurate in their social perceptions of others, then it would stand to reason that they would be more likely to make accurate perceptions in the current study. However, the gender and ethnicity of the target has been found to effect rater levels of accuracy in ratings, at least in ratings based upon visual behavior (Letzring, 2010). For example, female targets are more accurately perceived by others, regardless of the gender of the rater, but a gender and racial/ethnic match between raters and targets increases accuracy for female raters. However, the small sample sizes of the sample, coupled with the heterogeneity of the participants in terms of racial/ethnic groups, prevent an examination of the role that participant and target match

may play in the relationships being examined in the current study. Future directions may examine the effects of target and rater match in sociodemographics.

The post hoc analyses examining the level rater-target agreement on the individual personality factor level uncovered a different interesting pattern in the data. For example, while a main effect of self-description was found for all five factors, only Conscientiousness had significant interactions between ethnic group label accuracy and gender label accuracy, as well as one between gender label and self-description. This suggests that it is this factor that may have been especially influenced by the experimental manipulation. Conscientiousness is also among the more accurately judged traits in non-face-to-face interactions, such as in email address (Back et al., 2008) and in blog posts (Li & Chignell, 2010), as they may result in the a great deal of behavioral residue that can be utilized quite effectively by raters. This is likely due to conscientiousness being associated with having a strong sense of self-discipline and being organized (John et al., 2008; Li & Chignell, 2010). Non-face-to-face interactions, however, may be treated as more informal by some individuals than face-to-face interactions. As a result, the organization and self-discipline, whether through the use of punctuation, grammar, or some other linguistic behavior, that those who are higher in conscientiousness display during these interactions may stand out prominently, and thus would be able to provide evidence to their level on that personality trait (Li & Chignell, 2010). Future research examining the role of stereotypes and social perception may simply need to focus upon items assessing this factor to maximize effectiveness of any manipulations performed.



One findings of interest is the three-way Gender Label by Racial/Ethnic Label by Self-Description interaction on assessments of Openness to Experience. It may have resulted from the nature of the self-descriptions themselves. By expressing about a time that they felt happy, the authors of the self-descriptions may have expressed a greater deal of cues concerning Openness than other traits, as the prompt might have encouraged a greater deal of rumination about one's internal mental state at the time than might be typically expressed. Openness, by definition, deals with the complexity of one's internal states, thoughts, values, and beliefs (John et al., 2008). As a result, in scenarios where the internal states are more apparent than typical, such as when discussing a time when one felt a strong emotion, an observer would have access to cues concerning it and thus may be more attuned to perceive the individual on that trait. This is supported by the fact Vazire (2010) found that friends were much more accurate of their perceptions of an individual on Openness-related traits than strangers, but that they were equally accurate for a highly visible trait like Extraversion.

This study, while not supporting the hypotheses, falls within the realm of the current research. While previous research has found some support for the ability of strangers being able to judge the traits of others, there has been some variability in the as the findings concerning computer-mediated communication. The specific traits that can be accurately perceived by raters differ by the specific means of electronic communication being examined. For example, observer ratings based upon the content of personal emails have demonstrated high rater-target agreement for Extraversion (Gill et al., 2006) However, ratings of Extraversion based on email addresses demonstrated significant rater-target agreement (Back et al., 2008) have not, and the results have been

mixed for chat conversations (Gosling, Rentfrow & Swann, 2003; Markey & Wells, 2002). In addition, a meta-analysis of the rating of personality perception from text and online social network profiles found that while there is substantial evidence for the accurate perception of four out of the Five Factors, the evidence for Neuroticism is slightly weaker (Tskhay & Rule, 2014). Taken together, this suggests that while on a whole, the rating of personality based on written behavior can produce a level of accuracy; there remains some variability in the literature.

Another explanation for the non-significant findings touches on the purpose of person perception itself. Person perception is an inherently social process. It allows individuals to be able to navigate a complex social world by being able to infer information about others. This information, in turn, would impact how one would respond in order to achieve a goal or goals. This process requires at least some social context to the cues that lead to these judgments. If one examines the previous literature concerning social judgments from minimal information, the criteria used could have stronger social components than are present here. This is even true for studies that utilized static photographs of faces. The face is among the most salient parts of another person during interpersonal interactions. As a result, people may be especially skilled at making judgments based on this type of information as a result of experience (Rule, Tskhay, Freeman, Ambady, 2014). In the same way, much of the previous research utilizing electronic content used either stimuli that were inherently social, such as chat logs (Chan & Mendelsohn, 2010; Markey & Wells, 2002), social media content (Tskhay & Rule, 2014), and personal blog posts (Li & Chignell, 2010). Each one of these stimuli is designed to express some social message on the part of its creator. While one can argue

that there is some social element present in the current study in the self-descriptions that the targets provided, describing a particular event in their lives that evoked emotion. However, the event may not have been socially relevant to the target, and thus may not have provided enough useful information for the judges, a fact that may have been exasperated by the short length of the self-descriptions.

It could also be possible that the integrated nature of the study site, with a sizable presence of every ethnic group represented in the manipulation, means that our participants were less likely to draw upon stereotypes, and more likely to use idiosyncratic data, to form impressions of others. Judgments of others are heavily influenced by level of expertise and the ability to read cues, in order to achieve perceptive accuracy (Biesanz, 2010; Funder, 1995, 2012). The more that one knows and understands the meaning behind particular cues, the more effectively that one could use them in a judgment of another individual (Letzring, 2008).

Social group membership influences the formation of social judgments under SAM (Rogers & Biesanz, 2014). As a result, the more contact one has with members of various social groups, the greater level of perceptive accuracy that they may have when making assessments of members of those groups. According to the Contact Hypothesis prejudice can be reduced through having equal status cross-group relationships (Allport, 1954/1979). As the study site has a combination of having a relatively small, but diverse population, there is a good chance that participants have close relationships of some sort with a member or members of each of the target groups. Therefore, they may have been more likely to use their experience with these known individuals to influence their judgments of the study's targets than traditional stereotypes.

Future research in understanding of the use of stereotypes and prejudice in personality perception may examine the process through a more inherently social medium, such as online social media profiles or emails. Both formats have been used extensively in research on person perception and may provide additional information and cues that may be used in making judgments (Gill et al., 2006; Tskhay & Rule, 2014). However, no study is known to the author that integrates studies of prejudice in these domains. In addition, one of the limitations of this study was in the more explicit nature of some of the measures and manipulation used, which may have produced some level of unintended reactance in the participants, although we attempted to minimize this effect by administering the racism/sexism questionnaires after the profile ratings. Additional research could consider using implicit measures to either assess prejudice or manipulate group membership. This may be able to minimize such confounds that could bias results. Another avenue of future research could look at the extent that other forms of prejudice, such as class-based and/or sexuality-based, may influence individuals' perceptions of others based on textual information in general. Would believing a target belong to one of these invisible social groups produce a similar effect as have been shown in the literature for race and gender? Despite the inconsistent findings, the preceding study suggests several paths in which to follow for future research to better understand the processes underlying the social perception that allows humanity to navigate the complex social world in which they live.

Table 1

*Mean Level of Rater-Target Agreement by Self-Description*

<b>Self-description</b>	<b>Mean</b>	<b>N</b>	<b>Standard Deviation</b>
A	.1275942	104	.26659707
B	-.0397230	104	.26825198
C	.2702988	104	.28432322
D	.2981149	104	.34093253
E	.4034844	104	.22919525
F	.0885282	104	.16091847
G	-.2560744	104	.21650294
H	.1814195	104	.23413390
<b>Total</b>	<b>.1342053</b>	<b>832</b>	<b>.32035869</b>

Table 2

*Correlation Matrix of Prejudice Measures*

	MRS	CoBRAS	ISM-R	ASI	AMI	MSS	ISM-S	SDO
MRS		.605***	.593***	.188***	.037	.451***	.398***	.523***
CoBras			.507***	.109**	-.219***	.543***	.446***	.421***
ISM-R				.209**	.105**	.497***	.564***	.700***
ASI					.677***	.288***	.371***	.265***
AMI						.082*	.209***	.184***
MSS							.532***	.506***
ISM-S								.574***
SDO								

*Note.* MRS=Modern Sexism Scale; CoBRAS=Color-blind Racial Attitude Scale; ISM-R = Intolerant Schema Measure-Racism subscale; ASI=Ambivalent Sexism Inventory; AMI = Ambivalence Towards Men Inventory; MSS=Modern Sexism Scale; ISM-S= Intolerant Schema Measure-Sexism subscale; SDO = Social Dominance Orientation Scale. \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\* $p < .001$

Table 3

*Group Means of Rater-Target Accuracy for Separate Self-Descriptions*

Self- Description	Labels		Mean	Std. Deviation	N
	Gender Label	Racial/Ethnic Label			
A	Inaccurate	Inaccurate	.1953	.27786	19
		Accurate	.0575	.25329	23
		Total	.1198	.27044	42
	Accurate	Inaccurate	.0932	.27659	30
		Accurate	.1700	.25449	32
		Total	.1329	.26605	62
	Total	Inaccurate	.1328	.27874	49
		Accurate	.1230	.25779	55
		Total	.1276	.26659	104
	B	Inaccurate	Inaccurate	-.1287	.28970
Accurate			.0092	.26817	22
Total			-.0704	.28647	52
Accurate		Inaccurate	.0083	.25543	24
		Accurate	-.0239	.24451	28
		Total	-.0090	.24767	52

	Total	Inaccurate	-.0678	.28099	54
		Accurate	-.0093	.25306	50
		Total	-.0397	.26824	10
					4
C	Inaccurate	Inaccurate	.2170	.28622	29
		Accurate	.3472	.21830	23
		Total	.2746	.26418	52
	Accurate	Inaccurate	.3057	.31295	28
		Accurate	.2199	.29679	24
		Total	.2661	.30568	52
	Total	Inaccurate	.2606	.30030	57
		Accurate	.2822	.26640	47
		Total	.2703	.28433	10
					4
D	Inaccurate	Inaccurate	.3459	.32767	17
		Accurate	.2106	.38771	25
		Total	.2654	.36661	42
	Accurate	Inaccurate	.3568	.32637	32
		Accurate	.2813	.32142	30
		Total	.3203	.32356	62
	Total	Inaccurate	.3530	.32343	49



		Accurate	.2492	.35150	55
		Total			10
			.2981	.34093	4
E	Inaccurate	Inaccurate	.4555	.21506	32
		Accurate	.3565	.26394	26
		Total	.4111	.24119	58
	Accurate	Inaccurate	.3749	.19966	17
		Accurate	.4050	.22677	29
		Total	.3938	.21538	46
	Total	Inaccurate	.4275	.21133	49
		Accurate	.3821	.24395	55
		Total	.4035	.22920	10
					4
F	Inaccurate	Inaccurate	.1038	.14643	30
		Accurate	.1020	.21470	26
		Total	.1030	.17961	56
	Accurate	Inaccurate	.0512	.13920	19
		Accurate	.0851	.13441	29
		Total	.0717	.13589	48
	Total	Inaccurate	.0834	.14454	49
		Accurate	.0931	.17544	55

		Total	.0885	.16092	10
					4
G	Inaccurate	Inaccurate	-.2109	.21143	26
		Accurate	-.2397	.23803	30
		Total	-.2263	.22451	56
	Accurate	Inaccurate	-.2906	.21657	29
		Accurate	-.2909	.18785	19
		Total	-.2907	.20360	48
	Total	Inaccurate	-.2529	.21594	55
		Accurate	-.2595	.21931	49
		Total	-.2561	.21650	10
					4
H	Inaccurate	Inaccurate	.1412	.21529	28
		Accurate	.2140	.25321	30
		Total	.1789	.23648	58
	Accurate	Inaccurate	.1430	.22747	29
		Accurate	.2555	.23374	17
		Total	.1845	.23374	46
	Total	Inaccurate	.1421	.21959	57
		Accurate	.2290	.24460	47
		Total	.1814	.23415	10
					4

Table 4

*Group Means of Rater-Target Accuracy for Collapsed Across Self-Descriptions*

Self- Description	Labels		Mean	Std. Deviation	N
	Gender Label	Racial/Ethni c Label			
A	Inaccurate	Inaccurate	.1048244	.24623251	19
		Accurate	.0580441	.28026718	23
		Total	.0792066	.26329114	42
	Accurate	Inaccurate	.1713401	.29652335	30
		Accurate	.1500910	.23804379	32
		Total	.1603728	.26591810	62
	Total	Inaccurate	.1455483	.27736377	49
		Accurate	.1115986	.25813013	55
		Total	.1275942	.26659707	104
B	Inaccurate	Inaccurate	-.0498865	.27414650	30
		Accurate	.0388148	.27673993	22
		Total	-.0123590	.27609602	52
	Accurate	Inaccurate	-.1132240	.23975982	24
		Accurate	-.0275409	.27414586	28
		Total	-.0670870	.25994857	52
	Total	Inaccurate	-.0780365	.25899556	54

		Accurate	.0016556	.27448436	50
		Total	-.0397230	.26825198	104
C	Inaccurate	Inaccurate	.2149992	.34331160	29
		Accurate	.3682508	.23806179	23
		Total	.2827836	.30832339	52
	Accurate	Inaccurate	.2626369	.26549388	28
		Accurate	.2521876	.26022473	24
		Total	.2578141	.26054426	52
	Total	Inaccurate	.2384002	.30576710	57
		Accurate	.3089845	.25377638	47
		Total	.2702988	.28432322	104
D	Inaccurate	Inaccurate	.2871426	.34345460	17
		Accurate	.2508178	.30637944	25
		Total	.2655207	.31828694	42
	Accurate	Inaccurate	.1998510	.38091276	32
		Accurate	.4485617	.28088965	30
		Total	.3201949	.35629670	62
	Total	Inaccurate	.2301358	.36713709	49
		Accurate	.3586781	.30653770	55
		Total	.2981149	.34093253	104
E	Inaccurate	Inaccurate	.3676447	.23301736	32

		Accurate	.4994327	.20597309	26
		Total	.4267221	.22914695	58
	Accurate	Inaccurate	.4027335	.19770180	17
		Accurate	.3574494	.24636651	29
		Total	.3741848	.22836883	46
	Total	Inaccurate	.3798184	.21995525	49
		Accurate	.4245688	.23713133	55
		Total	.4034844	.22919525	104
F	Inaccurate	Inaccurate	.0927686	.18319289	30
		Accurate	.0368704	.15030808	26
		Total	.0668159	.16957507	56
	Accurate	Inaccurate	.0872824	.15717957	19
		Accurate	.1312717	.14160887	29
		Total	.1138593	.14792152	48
	Total	Inaccurate	.0906413	.17189368	49
		Accurate	.0866457	.15205194	55
		Total	.0885282	.16091847	104
G	Inaccurate	Inaccurate	-.2324373	.23369247	26
		Accurate	-.2174399	.20719066	30
		Total	-.2244030	.21798033	56
	Accurate	Inaccurate	-.2890948	.22816560	29

		Accurate	-.2990222	.18769345	19
		Total	-.2930244	.21102170	48
	Total	Inaccurate	-.2623113	.23041744	55
		Accurate	-.2490738	.20188968	49
		Total	-.2560744	.21650294	104
H	Inaccurate	Inaccurate	.1921480	.27019913	28
		Accurate	.1446549	.22247292	30
		Total	.1675826	.24563571	58
	Accurate	Inaccurate	.1794498	.21375491	29
		Accurate	.2319879	.23359682	17
		Total	.1988661	.22020267	46
	Total	Inaccurate	.1856875	.24101178	57
		Accurate	.1762434	.22799566	47
		Total	.1814195	.23413390	104

Table 5

*Gender Label Accuracy X Ethnicity Label Accuracy X Target ANOVA Neuroticism*

Source	df	F	Sig.	Partial $\eta^2$
Self-Description	7	25.858	.000*	.185
Ethnicity Label	1	.431	.511	.001
Gender Label	1	.350	.554	.000
Self-Description*	7	.797	.590	.007
Racial/Ethnic Label				
Self-Description * Gender	7	.836	.557	.007
Label				
Ethnicity Label* Gender Label	1	.003	.954	.000
Self-Description *				
Racial/Ethnic Label* Gender	7	1.375	.212	.012
Label				
Error	800			
Total	832			
Corrected Total	831			

Table 6

*Gender Label Accuracy X Ethnicity Label Accuracy X Target ANOVA Extraversion*

Source	df	F	Sig.	Partial $\eta^2$
Self-Description	7	12.320	.000*	.097
Ethnicity Label	1	.283	.595	.000
Gender Label	1	1.552	.213	.002
Self-Description*	7	.653	.712	.006
Racial/Ethnic Label				
Self-Description * Gender	7	1.004	.427	.009
Label				
Ethnicity Label* Gender	1	.325	.569	.000
Label				
Self-Description *				
Racial/Ethnic Label* Gender	7	.491	.841	.004
Label				
Error	800			
Total	832			
Corrected Total	831			



Table 7

*Gender Label Accuracy X Ethnicity Label Accuracy X Target ANOVA Openness*

Source	df	F	Sig.	Partial $\eta^2$
Self-Description	7	11.866	.000**	.094
Ethnicity Label	1	.425	.515	.001
Gender Label	1	.075	.784	.000
Self-Description*	7	.593	.762	.005
Racial/Ethnic Label				
Self-Description * Gender	7	.768	.614	.007
Label				
Ethnicity Label* Gender	1	1.472	.225	.002
Label				
Self-Description *				
Racial/Ethnic Label* Gender	7	3.911	.000**	.033
Label				
Error	800			
Total	832			
Corrected Total	831			

Table 8

*Gender Label Accuracy X Ethnicity Label Accuracy X Target ANOVA Agreeableness*

Source	df	F	Sig.	Partial $\eta^2$
Self-Description	7	5.502	.000**	.046
Ethnicity Label	1	2.820	.094	.004
Gender Label	1	.553	.457	.001
Self-Description*	7	1.066	.384	.009
Racial/Ethnic Label				
Self-Description * Gender	7	1.143	.334	.010
Label				
Ethnicity Label* Gender	1	.243	.622	.000
Label				
Self-Description *				
Racial/Ethnic Label* Gender	7	.674	.694	.006
Label				
Error	800			
Total	832			
Corrected Total	831			

Table 9

*Gender Label Accuracy X Ethnicity Label Accuracy X Target ANOVA Conscientiousness*

Source	df	F	Sig.	Partial $\eta^2$
Self-Description	7	9.015	.000***	.073
Ethnicity Label	1	.996	.319	.001
Gender Label	1	1.840	.175	.002
Self-Description*	7	1.271	.262	.011
Racial/Ethnic Label				
Self-Description * Gender	7	2.757	.008**	.024
Label				
Ethnicity Label* Gender	1	3.913	.048*	.005
Label				
Self-Description *				
Racial/Ethnic Label* Gender	7	.171	.991	.001
Label				
Error	800			
Total	832			
Corrected Total	831			

Table 10

*Correlation Matrix of Between Rater Personality Factors and Accuracy for Targets*

	A	B	C	D	E	F	G	H
N	-.167	-.225*	-.224*	-.121	-.233*	.114	.159	-.015
E	.086	.081	.078	.175	.206*	-.005	-.287**	-.023
O	.111	.065	.071	-.108	-.027	-.063	-.145	.133
A	.123	.241*	.379***	.281**	.165	-.262**	-.329**	-.002
C	.165	.155	.318**	.185	.179	-.117	-.268**	.054

*Note.* N= Rater Neuroticism; E= Rater Extraversion; O = Rater Openness; A= Rater Agreeableness; C= Rater Conscientiousness; A= Self-Description A; B= Self-Description B; C= Self-Description C; D= Self-Description D; E= Self-Description E; F= Self-Description F; G= Self-Description G; H= Self-Description H \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\* $p < .001$

Table 11

*Correlation Matrix between Rater Personality and Self-Other Agreement across Targets, Controlling for Prejudice*

		A	B	C	D	E	F	G	H
Prejudice	N	-.181	-.252**	-.245	-.128	-.267**	.103	.170	- .019
	E	.092	.092	.087	.178	.225*	.000	-.294**	- .021
	O	.091	.027	.040	-.122	-.077	-.088	-.129	.129
	A	.119	.237*	.378***	.279**	.160	-.269**	-.327***	- .004
	C	.193	.204*	.364***	.201*	.239*	-.096	-.294**	.062

*Note.* Prejudice= Global Prejudice Index; N= Rater Neuroticism; E= Rater Extraversion; O = Rater Openness; A= Rater Agreeableness; C= Rater Conscientiousness; A= Self-Description A; B= Self-Description B; C= Self-Description C; D= Self-Description D; E= Self-Description E; F= Self-Description F; G= Self-Description G; H= Self-Description H \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\* $p < .001$

Table 12

*Correlation Matrix of Between Rater Personality and General Accuracy for Targets*

	N	E	O	A	C
Accuracy	-.083*	.041	.005	.088*	.084*

*Note.* Accuracy= Rater-Agreement N= Rater Neuroticism; E= Rater Extraversion; O = Rater Openness; A= Rater Agreeableness; C= Rater Conscientiousness

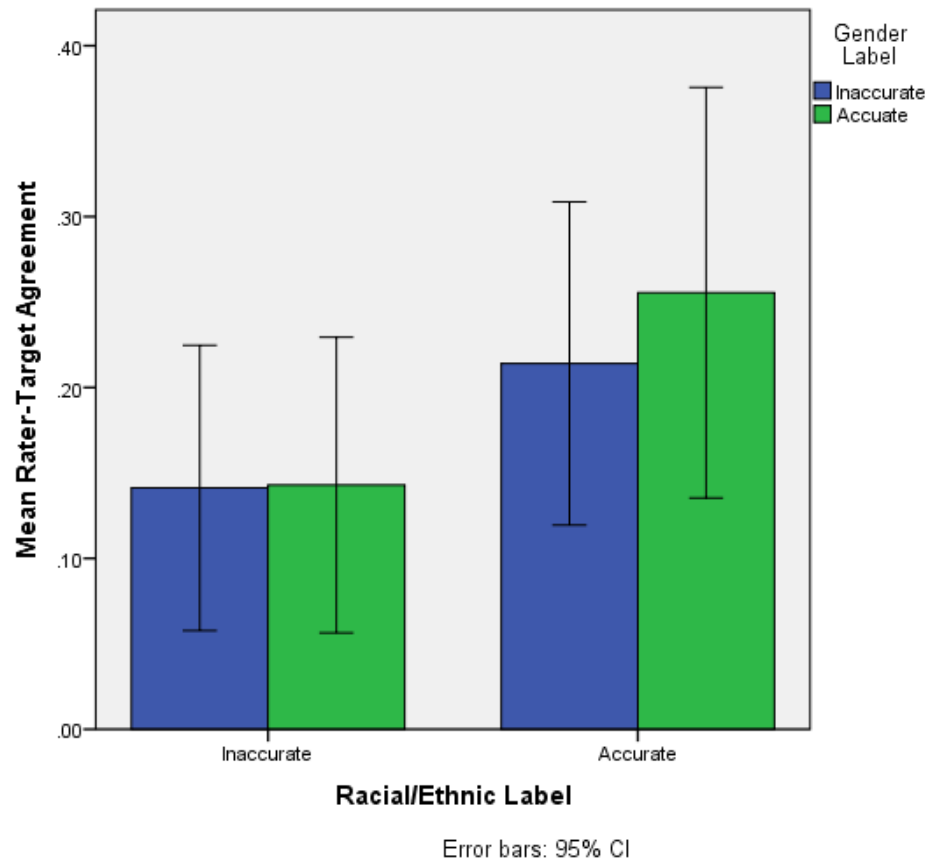
\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

Table 13

*Correlation Matrix of Between Rater Personality and General Accuracy Controlling for Prejudice*

		N	E	O	A	C
Prejudice	Accuracy	-.092**	.045	-.011	.084*	.103*

*Note.* Prejudice= General Prejudice Index; Accuracy= Rater-Agreement N= Rater Neuroticism; E= Rater Extraversion; O = Rater Openness; A= Rater Agreeableness; C= Rater Conscientiousness \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\* $p < .001$



*Figure 1.* Effects of racial/ethnic label and gender label accuracy on mean rater-target agreement for self-description H.



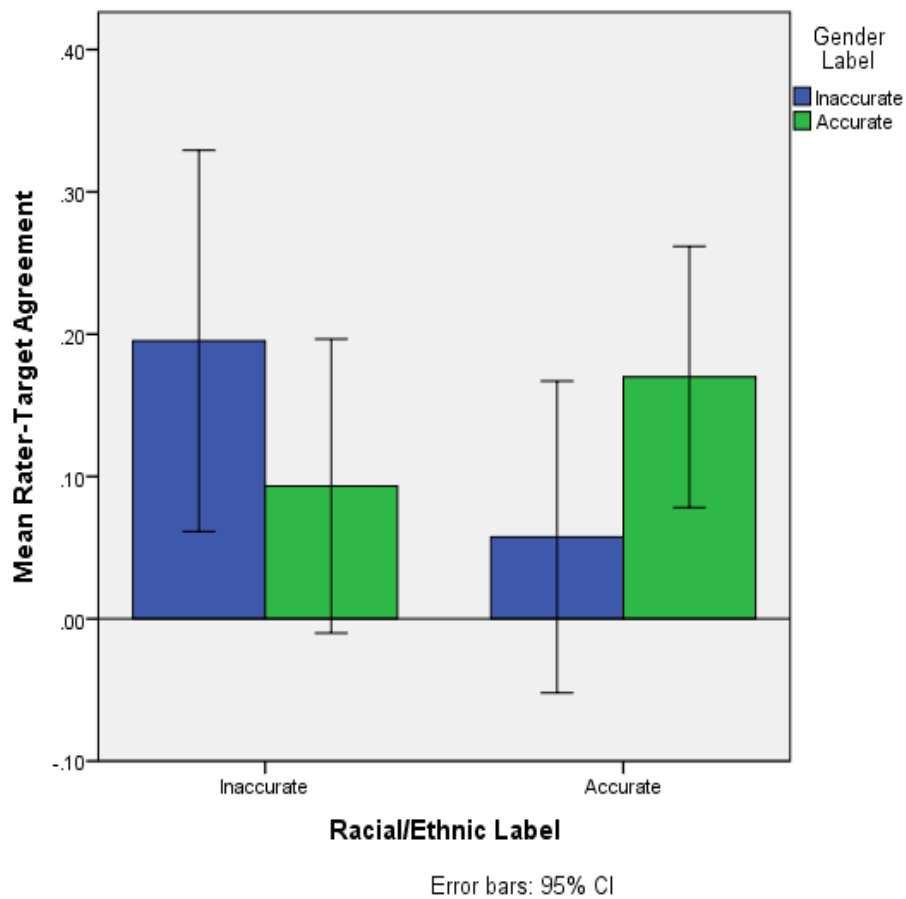
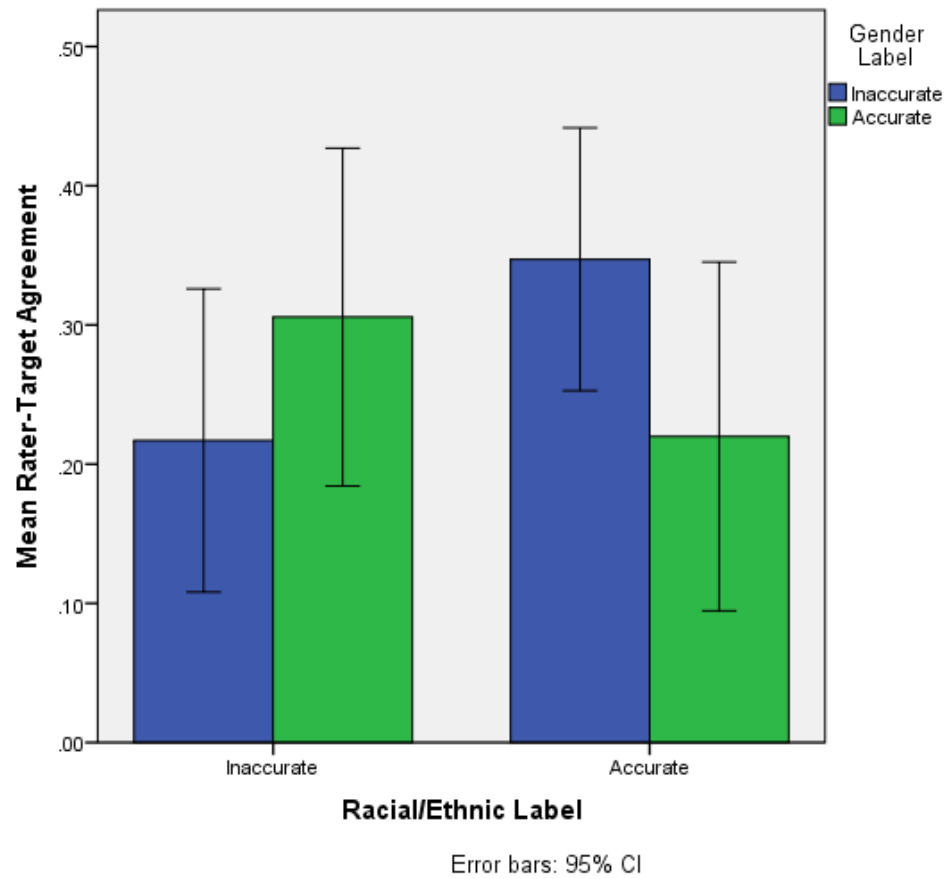


Figure 2. Effects of racial/ethnic label and gender label accuracy on mean rater-target agreement for self-description A.



*Figure 3.* Effects of racial/ethnic label and gender label accuracy on mean rater-target agreement for self-description C.

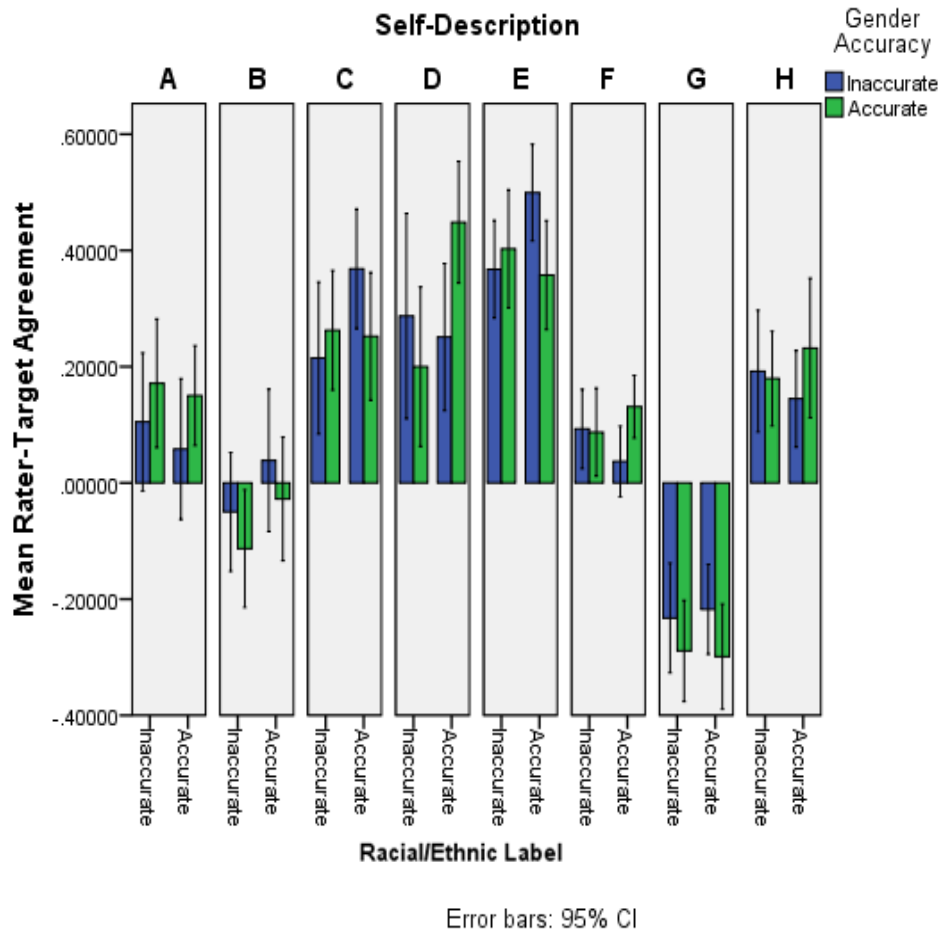
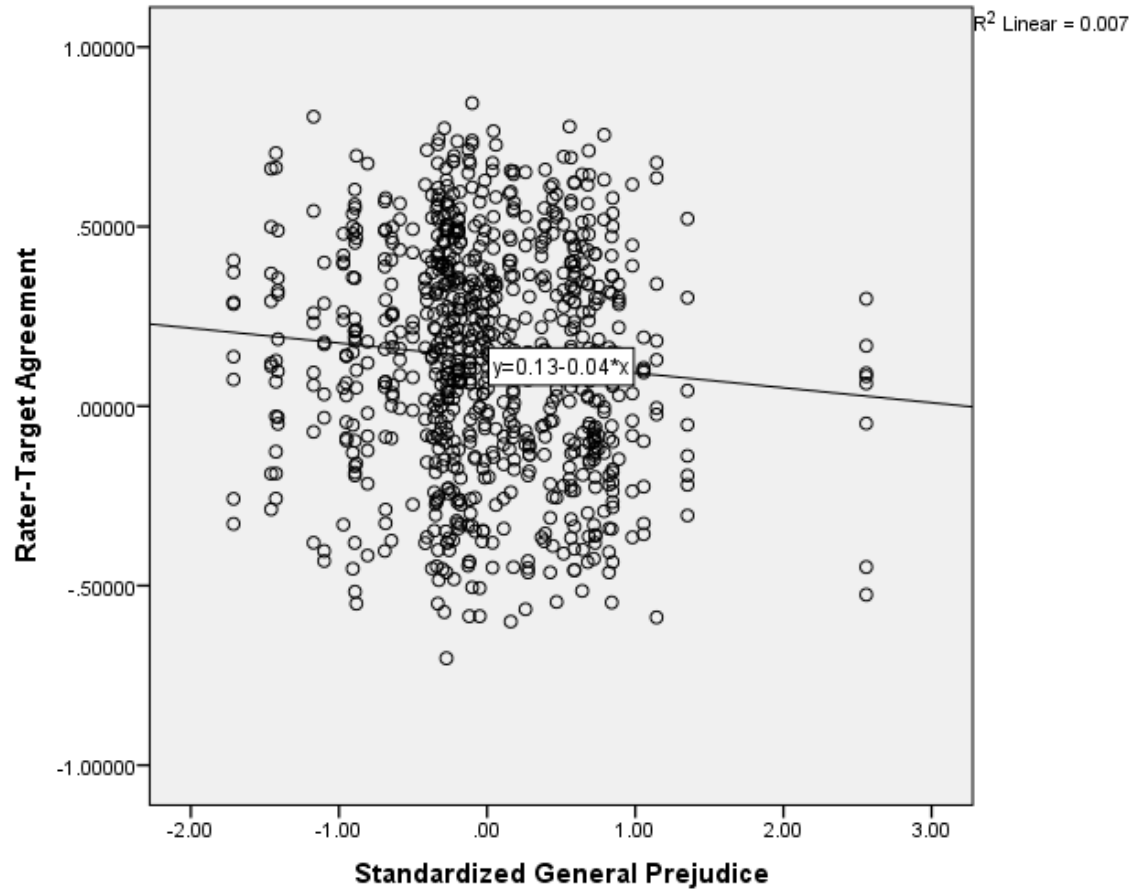
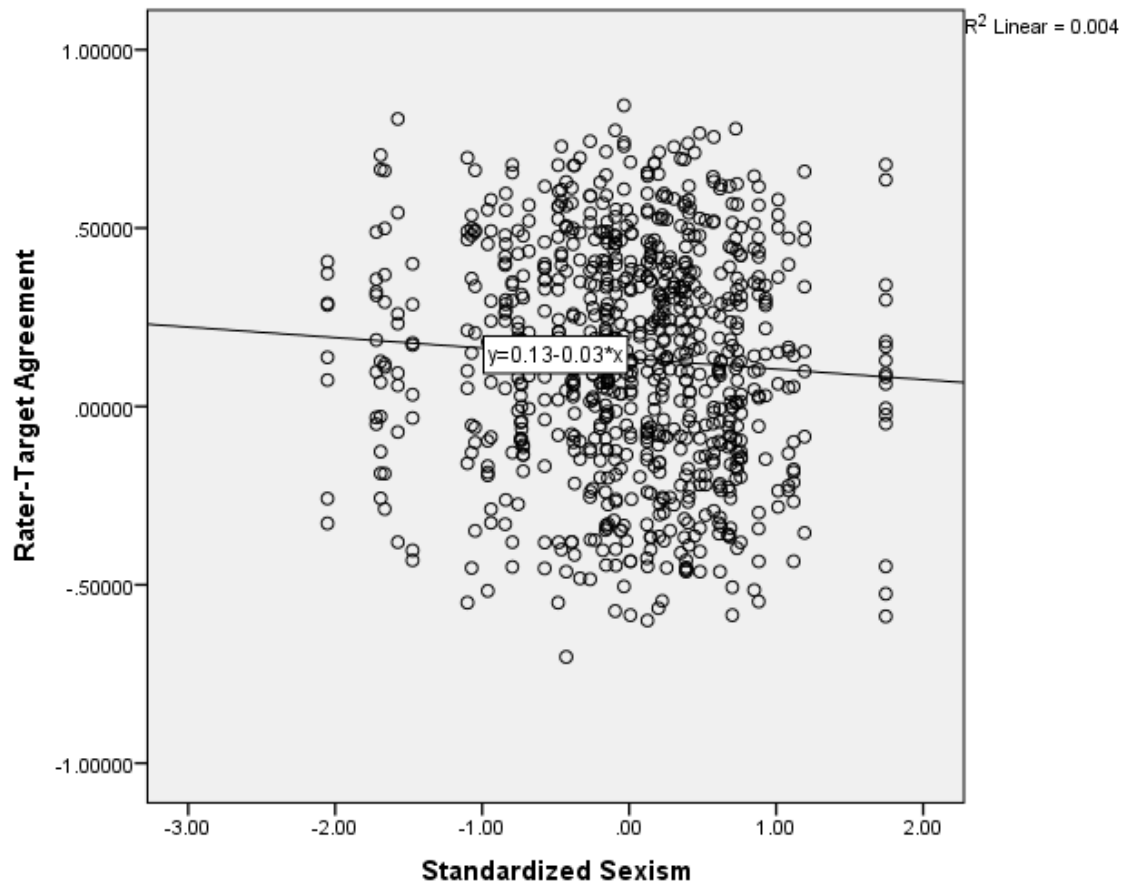


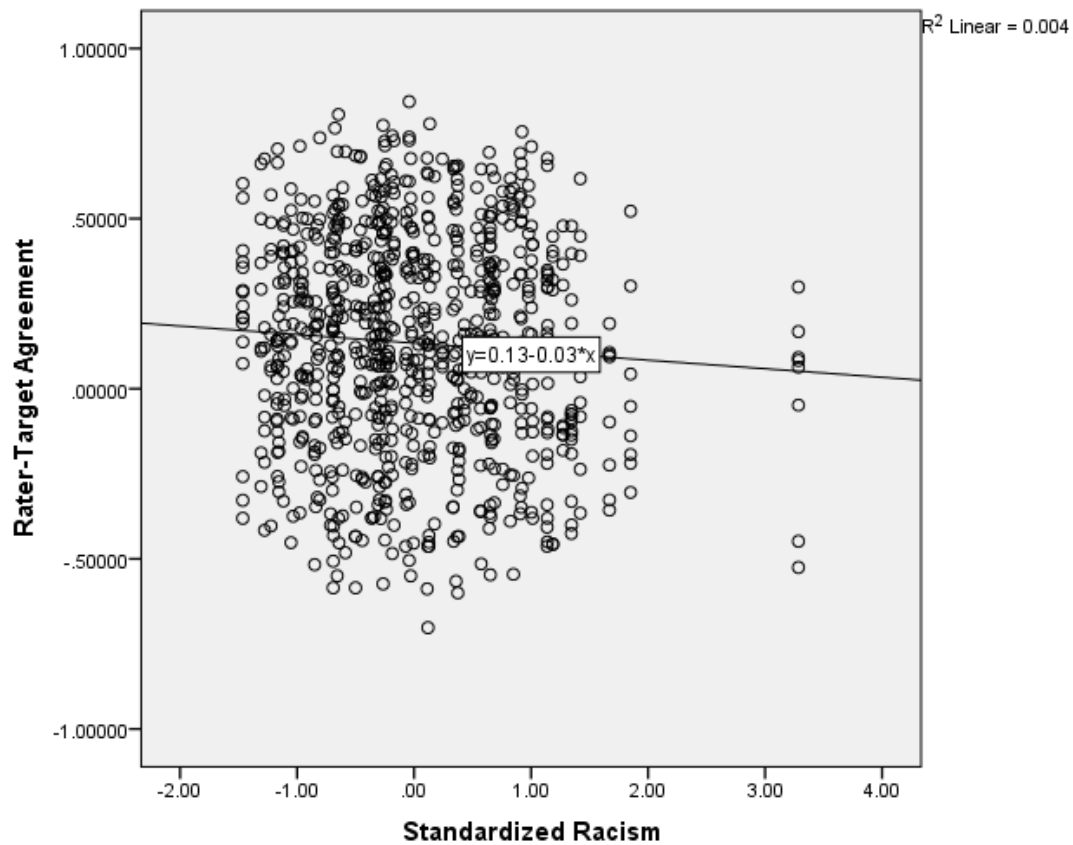
Figure 4. A comparison of the effects of racial/ethnic label accuracy and gender label accuracy on mean rater-target agreement by different self-descriptions.



*Figure 5.* Relationship between standardized index of global prejudice and rater-target agreement with fit line.



*Figure 5.* Relationship between standardized index of sexism and rater-target agreement with fit line.



*Figure 6.* Relationship between standardized index of racism and rater-target agreement with fit line.

## Appendices

### Appendix A

Alphabetical List of Trait Terms Selected from the Adjective Check List (Gough & Heilbrun, 1983) Used in Williams and Mendelsohn (2008).

<i>Affectionate</i>	<i>Feminine</i>	<i>Self-centered</i>
<i>Ambitious</i>	<i>Flirtatious</i>	<i>Self-controlled</i>
<i>Argumentative</i>	<i>Good-natured</i>	<i>Shy</i>
<i>Boastful</i>	<i>Helpful</i>	<i>Submissive</i>
<i>Capable</i>	<i>Immature</i>	<i>Sympathetic</i>
<i>Coarse</i>	<i>Inventive</i>	<i>Unemotional</i>
<i>Cold</i>	<i>Irritable</i>	<i>Unselfish</i>
<i>Confident</i>	<i>Logical</i>	<i>Warm</i>
<i>Dependent</i>	<i>Loyal</i>	<i>Worrying</i>
<i>Egotistical</i>	<i>Masculine</i>	
<i>Enterprising</i>	<i>Moody</i>	

## Appendix B

## Big Five Inventory

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 select a number for each statement indicating the extent to which you agree or disagree with the characteristic.

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 C

## National Character Survey

Which of these sets of adjectives best describe [you/the author]?

- |                                       |       |                                 |
|---------------------------------------|-------|---------------------------------|
| 1. Anxious, nervous, worrying         | ----- | At ease, calm, relaxed          |
| 2. Friendly, warm, affectionate       | ----- | Cold, aloof, reserved           |
| 3. Imaginative, a dreamer             | ----- | Practical, down-to-earth        |
| 4. Trusting, gullible, naïve          | ----- | Suspicious, skeptical, cynical  |
| 5. Capable, efficient, competent      | ----- | Inept, unprepared               |
| 6. Even-tempered, easy-going          | ----- | Irritable, angry, touchy        |
| 7. Solitary, shy, avoids crowds       | ----- | Gregarious, sociable, outgoing  |
| 8. Unartistic, uninterested in art    | ----- | Sensitive to art and beauty     |
| 9. Crafty, sly, manipulative          | ----- | Frank, sincere, straightforward |
| 10. Disorganized, sloppy              | ----- | Organized, neat, methodical     |
| 11. Depressed, sad, pessimistic       | ----- | Contented, optimistic           |
| 12. Assertive, forceful, dominant     | ----- | Submissive, a follower          |
| 13. Emotionally sensitive, passionate | ----- | Unfeeling, unempathic           |
| 14. Generous, giving, considerate     | ----- | Selfish, stingy, greedy         |
| 15. Dutiful, scrupulous               | ----- | Unreliable, undependable        |
| 16. Poised, comfortable with others   | ----- | Self-conscious, awkward, timid  |
| 17. Slow, lethargic, unenergetic      | ----- | Active, vigorous, busy          |
| 18. Habit-bound, prefers routine      | ----- | Innovative, prefers variety     |
| 19. Aggressive, competitive, stubborn | ----- | Compliant, cooperative, docile  |

20. Lazy, unambitious, aimless	-----	Ambitious, workaholic
21. Impulsive, yielding to temptation	-----	Controlled, self-restrained
22. Adventurous, fun-loving, risk-taking	-----	Avoids excitement, stimulation
23. Intellectually curious, open-minded	-----	Narrow interests, bored by ideas
24. Modest, humble, self-effacing	-----	Arrogant, conceited
25. Disciplined, persistent, strong-willed	-----	Procrastinating, quitting, weak
26. Resilient, copes well with crises	-----	Vulnerable, fragile, helpless
27. Somber, dull, sober	-----	Happy, cheerful, joyous
28. Dogmatic, traditional, conservative	-----	Liberal, free-thinking
29. Ruthless, hard-headed, unsentimental	-----	Sympathetic, humanitarian
30. Spontaneous, careless, thoughtless	-----	Cautious, reflective, careful

## Appendix D

## Modern Racism Scale (McConahay, 1986)

Directions: Below are a number of opinion statements about public issues, policies, and beliefs about your world in general. You will agree with some, disagree with some and have no opinion about others. You are under no obligation to give an opinion on any item. Using the 5-point scale, please give your honest rating about the degree to which you personally agree or disagree with each statement.

1	2	3	4	5
Strongly Agree			Strongly Disagree	

1. \_\_\_Discrimination against blacks is no longer a problem in the United States. (R)
2. \_\_\_It is easy to understand the anger of black people in America.
3. \_\_\_ Blacks have more influence upon school desegregation plans than they ought to have. (R)
4. \_\_\_Blacks are getting too demanding in their push for equal rights. (R)
5. \_\_\_Blacks should not push themselves where they are not wanted. (R)
6. \_\_\_Over the past few years, blacks have gotten more economically than they deserve. (R)
7. \_\_\_Over the past few years, the government and news media have shown more respect to blacks than they deserve. (R)

(R) = reverse-scored item.

## Appendix E

## Color-blind Racial Attitude Scale (CoBRAS)

Directions. The following is a set of questions that deal with social issues in the United States (U.S.). Using the 6-point scale, please give your honest rating about the degree to which you personally agree or disagree with each statement. Please be as open and honest as you can; there are no right or wrong answers

1                      2                      3                      4                      5                      6

Strongly Disagree

Strongly Agree

1. \_\_\_White people in the U.S. have certain advantages because of the color of their skin. (R)
2. \_\_\_Race is very important in determining who is successful and who is not. (R)
3. \_\_\_Race plays an important role in who gets sent to prison. (R)
4. \_\_\_Race plays a major role in the type of social services (such as type of health care or day v care) that people receive in the U.S. (R)
5. \_\_\_Racial and ethnic minorities do not have the same opportunities as white people in the U.S. (R)
6. \_\_\_Everyone who works hard, no matter what race they are, has an equal chance to become rich.
7. \_\_\_White people are more to blame for racial discrimination than racial and ethnic minorities.(R)
8. \_\_\_Social policies, such as affirmative action, discriminate unfairly against White people.

9. \_\_\_ White people in the U.S. are discriminated against because of the color their skin.
10. \_\_\_ English should be the only official language in the U.S.
11. \_\_\_ Due to racial discrimination, programs such as affirmative action are necessary to help create equality. (R)
12. \_\_\_ Racial and ethnic minorities in the U.S. have certain advantages because of the color of their skin.
13. \_\_\_ It is important that people begin to think of themselves as American and not African American, Mexican American or Italian American.
14. \_\_\_ Immigrants should try to fit into the culture and adopt the values of the U.S.
15. \_\_\_ Racial problems in the U.S. are rare, isolated situations.
16. \_\_\_ Talking about racial issues causes unnecessary tension.
17. \_\_\_ Racism is a major problem in the U.S. (R)
18. \_\_\_ It is important for public schools to teach about the history and contributions of racial and ethnic minorities. (R)
19. \_\_\_ It is important for political leaders to talk about racism to help work through or solve society's problems. (R)
20. \_\_\_ Racism may have been a problem in the past, but it is not an important problem today.

(R) = reverse-scored item.



## Appendix F

## The Intolerant Schema Measure (ISM)-Racism subscale

*Intolerant Schema Measure (ISM)-Racism*

Instructions: Please indicate how descriptive each statement is of your beliefs by circling the number that corresponds to your response. (1 = *strongly disagree* to 5 = *strongly agree*)

1. \_\_\_ I favor laws that permit racial minority persons to rent or purchase houses, even when the person offering the property for sale or rent does not wish to sell or rent to minorities. (R)
2. \_\_\_ Racial minorities have more influence on school desegregation plans than they ought to have.
3. \_\_\_ Racial minorities are getting too demanding in their push for equal rights.
4. \_\_\_ It is a bad idea for racial minorities and Whites to marry one another.
5. \_\_\_ Racial minorities should not push themselves where they are not wanted.
6. \_\_\_ If a racial minority family with about the same income and education as I have moved in next door, I would mind a great deal.
7. \_\_\_ It was wrong for the United States Supreme Court to outlaw segregation in its 1954 decision.

8.\_\_\_\_ Over the past few years, racial minorities have gotten more economically than they deserve.

9.\_\_\_\_ Over the past few years, the government and news media have shown more respect to racial minorities than they deserve.

*Note.*

Subscale scores are calculated by averaging the 9 items (resulting in a range from 1 to 5, with higher scores indicating higher intolerance).

(R) = reverse-scored item.

## Appendix G

## The Ambivalent Sexism Inventory

**Relationships Between Men and Women**

Below is a series of statements concerning men and women and their relationships in contemporary society. Please indicate the degree to which you agree or disagree with each statement using the following scale: 0 = disagree strongly; 1 = disagree somewhat; 2 = disagree slightly; 3 = agree slightly; 4 = agree somewhat; 5 = agree strongly.

- \_\_\_ 1. No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman.
- \_\_\_ 2. Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for "equality."
- \_\_\_ 3. In a disaster, women ought not necessarily to be rescued before men. (R)
- \_\_\_ 4. Most women interpret innocent remarks or acts as being sexist.
- \_\_\_ 5. Women are too easily offended.
- \_\_\_ 6. People are often truly happy in life without being romantically involved with a member of the other sex. (R)
- \_\_\_ 7. Feminists are not seeking for women to have more power than men. (R)
- \_\_\_ 8. Many women have a quality of purity that few men possess.
- \_\_\_ 9. Women should be cherished and protected by men.

- \_\_\_ 10. Most women fail to appreciate fully all that men do for them.
- \_\_\_ 11. Women seek to gain power by getting control over men.
- \_\_\_ 12. Every man ought to have a woman whom he adores.
- \_\_\_ 13. Men are complete without women. (R)
- \_\_\_ 14. Women exaggerate problems they have at work.
- \_\_\_ 15. Once a woman gets a man to commit to her, she usually tries to put him on a tight leash.
- \_\_\_ 16. When women lose to men in a fair competition, they typically complain about being discriminated against.
- \_\_\_ 17. A good woman should be set on a pedestal by her man.
- \_\_\_ 18. There are actually very few women who get a kick out of teasing men by seeming sexually available and then refusing male advances. (R)
- \_\_\_ 19. Women, compared to men, tend to have a superior moral sensibility.
- \_\_\_ 20. Men should be willing to sacrifice their own well-being in order to provide financially for the women in their lives.
- \_\_\_ 21. Feminists are making entirely reasonable demands of men. (R)
- \_\_\_ 22. Women, as compared to men, tend to have a more refined sense of culture and good taste.

## Appendix H

## Ambivalence Toward Men Inventory

Below are a series of statements concerning men and women and their relationships in contemporary society. Please indicate the degree to which you agree or disagree with each statement using the scale below:

0	1	2	3	4	5
Disagree	Disagree	Disagree	Agree	Agree	Agree
Strongly	Somewhat	Slightly	Slightly	Somewhat	Strongly

- \_\_\_ 1. Even if both members of a couple work, the woman ought to be more attentive to taking care of her man at home.
- \_\_\_ 2. A man who is sexually attracted to a woman typically has no morals about doing whatever it takes to get her to bed.
- \_\_\_ 3. Men are less likely to fall apart in emergencies than women are.
- \_\_\_ 4. Men who act to “help” women, they are often trying to prove they are better than women.
- \_\_\_ 5. Every woman needs a male partner who will cherish her.
- \_\_\_ 6. Men would be lost in this world if women weren’t there to guide them.
- \_\_\_ 7. A woman will never be truly fulfilled in life if she doesn’t have a committed, long-term relationship with a man.

- \_\_\_ 8. Men act as babies when they are sick.
- \_\_\_ 9. Men will always fight to have greater control in society than women.
- \_\_\_ 10. Men are mainly useful to provide financial security for women.
- \_\_\_ 11. Even men who claim to be sensitive to women's rights really want a traditional relationship at home, with the woman performing most of the housekeeping and child care.
- \_\_\_ 12. Even woman ought to have a man she adores.
- \_\_\_ 13. Men are more willing to put themselves in danger to protect others.
- \_\_\_ 14. Men usually try to dominate conversations when talking to women.
- \_\_\_ 15. Most men pay lip service to equality for women, but can't handle having a woman as an equal.
- \_\_\_ 16. Women are incomplete without men.
- \_\_\_ 17. When it comes down to it, most men are really like children.
- \_\_\_ 18. Men are more willing to take risks than women.
- \_\_\_ 19. Most men sexually harass women, even if only in subtle ways, once they are in a position of power over them.
- \_\_\_ 20. Women ought to take care of their men at home, because men would fall apart if they had to fend for themselves.

## Appendix I

## Modern Sexism Scale

Directions: Below are a number of opinion statements about public issues, policies, and beliefs about your world in general. You will agree with some, disagree with some and have no opinion about others. You are under no obligation to give an opinion on any item. Using the 5-point scale, please give your honest rating about the degree to which you personally agree or disagree with each statement.

1	2	3	4	5
Strongly Agree			Strongly Disagree	

1. \_\_\_ Discrimination against women is no longer a problem in the United States. (R)
2. \_\_\_ Women often miss out on good jobs due to sexual discrimination.
3. \_\_\_ It is rare to see women treated in a sexist manner on television. (R)
4. \_\_\_ On average, people in our society treat husbands and wives equally. (R)
5. \_\_\_ Society has reached the point where women and men have equal opportunities for achievement. (R)
6. \_\_\_ It is easy to understand the anger of women's groups in America
7. \_\_\_ It is easy to understand why women's groups are still concerned about societal limitations of women's opportunities.

8. \_\_\_Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences.

(R)

(R) = reverse-scored item.



## Appendix J

## The Intolerant Schema Measure (ISM)-Sexism

*Intolerant Schema Measure (ISM)-Sexism*

Instructions: Please indicate how descriptive each statement is of your beliefs by selecting the number that corresponds to your response. (1 = *strongly disagree* to 5 = *strongly agree*)

1. \_\_\_ Women should worry less about their rights and more about becoming good wives and mothers.
2. \_\_\_ It is ridiculous for a woman to run a locomotive and for a man to darn socks.
3. \_\_\_ The intellectual leadership of a community should be largely in the hands of men.
4. \_\_\_ In general, the father should have greater authority than the mother in bringing up the children.
5. \_\_\_ There are many jobs in which men should be given preference over women in being hired or promoted.
6. \_\_\_ Women shouldn't push themselves where they are not wanted.
7. \_\_\_ Women's requests in terms of equality between the sexes are simply exaggerated.
8. \_\_\_ Over the past few years, women have gotten more from government than they deserve.

9. \_\_\_Universities are wrong to admit women in costly programs such as medicine, when in fact, a large number will leave their jobs after a few years to raise their children.

*Note.*

Subscale scores are calculated by averaging the 9 items (resulting in a range from 1 to 5, with higher scores indicating higher intolerance).

## Appendix K

## Social Dominance Orientation (SDO) Scale

Directions: Which of the following objects or statements do you have a positive or negative feeling toward? Beside each object or statement, fill in the circle for the number, from '1' to '7', which represents the degree of your positive or negative feeling.

1	2	3	4	5	6	
						7
Very Negative						Very
Positive						

1. \_\_\_ Some groups of people are simply inferior to other groups.
2. \_\_\_ In getting what you want, it is sometimes necessary to use force against other groups.
3. \_\_\_ It's OK if some groups have more of a chance in life than others.
4. \_\_\_ To get ahead in life, it is sometimes necessary to step on other groups.
5. \_\_\_ If certain groups stayed in their place, we would have fewer problems
6. \_\_\_ It's probably a good thing that certain groups are at the top and other groups are at the bottom.
7. \_\_\_ Inferior groups should stay in their place.
8. \_\_\_ Sometimes other groups must be kept in their place.

9. \_\_\_It would be good if groups could be equal. (R)
10. \_\_\_Group equality should be our ideal. (R)
11. \_\_\_All groups should be given an equal chance in life. (R)
12. \_\_\_We should do what we can to equalize conditions for different groups. (R)
13. \_\_\_Increased social equality. (R)
14. \_\_\_We would have fewer problems if we treated people more equally. (R)
15. \_\_\_We should strive to make incomes as equal as possible. (R)
16. \_\_\_No one group should dominate in society. (R)

(R) = reverse-scored item.

## Appendix L

## Personality Self-Description Stimuli

- A. My dad travels for a living and sometimes its for a couple weeks at a time. I am always very happy when he returns home from a long trip for work.
- B. One time I was really happy was I got accepted into this school. It was the middle of summer and I didn't get accepted into New Brunswick or Newark. the day i was going to accept for county college I got a call from this school.
- C. A couple of days ago watching my son being inducted into the Junior Honor Society, it was proud moment for me listening to moderator call his name and hear about the the things that he has accomplished thus far in his life.
- D. I felt really happy last week when I found out I got an internship.
- E. A time when I felt happy was when I earned my Associate's degree from CCC in 2012. Although I was disappointed that my family couldn't be there, I felt proud, excited, and the need to tell my kids that they can keep going forward to accomplish any goals that they wish to pursue!
- F. Whenever I make a joke without really thinking about it. I love realizing how my mind works without me putting conscious effort into making the joke and then thinking about it and realizing how funny it was.
- G. I was happy when I did a play in High School in which I got one of the leading roles. It felt good to be doing something productive in which I had a big part in. I felt competent, needed and fulfilled. I also enjoy theatre a great deal and was around good friends.
- H. winning semi finals at the final four last year with RUCMS

## References

- Allik, J., Mõttus, R., & Realo, A. (2010). Does national character reflect mean personality traits when both are measured by the same instrument?. *Journal of Research in Personality, 44*(1), 62-69. doi: 10.1016/j.jrp.2009.10.008
- Allport, G. W. (1979). *The nature of prejudice* (25th Anniversary ed.). Reading, MA: Addison-Wesley Publishing Company (Original work published 1954).
- Ambady, N. (2010). The perils of pondering: Intuition and thin slice judgments. *Psychological Inquiry, 21*(4), 271-278. doi:10.1080/1047840x.2010.524882
- Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83*(4), 947-961. doi:10.1037/0022-3514.83.4.947
- Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*(3), 518-529. doi:10.1037/0022-3514.69.3.518
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256-274. doi: 10.1037/0033-2909.111.2.256
- Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9*, 272-279. doi: 10.1037/h0025907
- Aosved, A. C., Long, P. J., & Voller, E. K. (2009). Measuring sexism, racism, sexual prejudice, ageism, classism, and religious intolerance: The intolerant schema measure. *Journal of Applied Social Psychology, 39*(10), 2321-2354. doi: 10.1111/j.1559-1816.2009.00528.x
- Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality, 42*(4), 1116-1122. doi: 10.1016/j.jrp.2008.02.001
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?. *Perspectives on Psychological Science, 2*(4), 396-403. doi: 10.1111/j.1745-6916.2007.00051.x
- Beer, A., & Watson, D. (2010). The effects of information and exposure on self-other agreement. *Journal of Research in Personality, 44*(1), 38-45. doi: 10.1016/j.jrp.2009.10.002
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*(2), 155-162. doi: 10.1037/h0036215

- Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, *45*(5), 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: 2010-09433-02
- 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*(1), 3-5. doi: 10.1177/1745691610393980
- Carlson, E. N., Vazire, S., & Oltmanns, T. F. (2013, April 20). Self-other knowledge asymmetries in personality pathology. *Journal of Personality*, *81*(2), 155-170. doi:10.1111/j.1467-6494.2012.00794.x
- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*(5), 1054-1072. doi:10.1016/j.jrp.2007.01.004
- Chan, W., & Mendelsohn, G. A. (2010). Disentangling stereotype and person effects: Do social stereotypes bias observer judgment of personality?. *Journal of Research in Personality*, *44*(2), 251-257. doi: 10.1016/j.jrp.2010.02.001
- Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., Löckenhoff, C. E., De Bolle, M. ... & Terracciano, A. (2012). Stereotypes of age differences in personality traits: Universal and accurate?. *Journal of Personality and Social Psychology*, *103*(6), 1050-1066. doi: 10.1037/a0029712
- Cho, J. C., & Knowles, E. D. (2013). I'm like you and you're like me: Social projection and self-stereotyping both help explain self-other correspondence. *Journal of Personality and Social Psychology*, *104*(3), 444-456. doi: 10.1037/a0031017
- Crawford, J. T., Jussim, L., Madon, S., Cain, T. R., & Stevens, S. T. (2011). The use of stereotypes and individuating information in political person perception. *Personality and Social Psychology Bulletin*, *37*(4), 529-542. doi: 10.1177/0146167211399473
- Darley, J., & Gross, P. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20. doi:10.1037/0022-3514.44.1.20
- Diekmann, A. B., Eagly, A. H., & Kulesa, P. (2002). Accuracy and bias in stereotypes about the social and political attitudes of women and men. *Journal of Experimental Social Psychology*, *38*(3), 268-282. doi: 10.1006/jesp.2001.1511
- Dovidio, J. F., Gaertner, S. L., & Kawakami, (2010). Racism. In J. Dovidio, M. Hewstone, P. Glick & V. Esses (Eds.), *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (pp. 312-327). London, England: SAGE.

- Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In J. Dovidio, M. Hewstone, P. Glick & V. Esses (Eds.), *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (pp. 3-28). London, England: SAGE.
- Funder, D. C. (1991). Global traits: A Neo-Allportian approach to personality. *Psychological Science*, 2(1), 31-39. doi:10.1111/j.1467-9280.1991.tb00093.x
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652-670. doi:10.1037/0033-295x.102.4.652
- Funder, D. C. (2001). Personality. *Annual Review of Psychology*, 52, 197-221. doi: 10.1146/annurev.psych.52.1.19
- Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality*, 43(2), 120-126. doi: 10.1016/j.jrp.2008.12.041
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177-182. doi: 10.1177/0963721412445309
- Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, 23(5), 369-401. doi: 10.1002/per.724
- Gamst, G. C., Liang, C. T. H., & Der-Karabetian, A. (2011). Racism- and prejudice-related measures. In *Handbook of Multicultural Measures*. Thousand Oaks, CA: SAGE Publications, Inc.
- Gilbert, G. M. (1951). Stereotype persistence and change among college students. *The Journal of Abnormal and Social Psychology*, 46(2), 245-254. doi: 10.1037/h0053696
- Gill, A. J., Oberlander, J., & Austin, E. (2006). Rating e-mail personality at zero acquaintance. *Personality and Individual Differences*, 40(3), 497-507. doi: 10.1016/j.paid.2005.06.027
- Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491-512. doi:10.1037/0022-3514.70.3.491
- Glick, P., & Fiske, S. T. (1999). The ambivalence toward men inventory: Differentiating hostile and benevolent beliefs about men. *Psychology of Women Quarterly*, 23(3), 519-536. doi:10.1111/j.1471-6402.1999.tb00379.x
- Glick, P., & Rudman, L. A. (2010). Sexism. In J. Dovidio, M. Hewstone, P. Glick & V. Esses (Eds.), *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (pp. 328-344). London, England: SAGE.
- Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216-1229. doi: 10.1037/0022-3514.59.6.1216



- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82(3), 379-398. doi: 10.1037//0022-3514.82.3.379
- Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, 66, 877-902. doi: 10.1146/annurev-psych-010814-015321
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504-528. doi:10.1016/s0092-6566(03)00046-1
- Gosling, S. D., Sandy, C. J., John, O. P., & Potter, J. (2010). Wired but not WEIRD: The promise of the Internet in reaching more diverse samples. *Behavioral and Brain Sciences*, 33, 94-95. doi:10.1017/s0140525x10000300
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93-104. doi: 10.1037/0003-066X.59.2.93
- Gough, H. G., & Heilbrun, A. B. (1983). *The adjective check list manual*. Palo Alto, CA: Consulting Psychologists Press.
- Hall, J. A., Schmid Mast, M., & Latu, I. M. (2015). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior*, 39(2), 131-163. doi: DOI 10.1007/s10919-014-0205-1
- Hancock, J. T., & Dunham, P. J. (2001). Impression formation in computer-mediated communication revisited: An analysis of the breadth and intensity of impressions. *Communication Research*, 28(3), 325-347. doi:10.1177/009365001028003004
- Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, 19(2), 49-71. doi: 10.1080/10478400802049936
- Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47, 237-271. doi: 10.1146/annurev.psych.47.1.237
- Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality*, 43(3), 524-527. doi: 10.1016/j.jrp.2009.01.006
- Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of Research in Personality*, 45(1), 92-99. doi: 10.1016/j.jrp.2010.11.015
- Hřebíčková, M. & Graf, S. (2014), Accuracy of national stereotypes in Central Europe: Outgroups are not better than ingroup in considering personality traits of real people. *European Journal of Personality*, 28 (1): 60-72. doi: 10.1002/per.1904

- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five Trait Taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.
- Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York, NY: Oxford University Press.
- Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. Nelson (Ed.), *Handbook of Prejudice, Stereotyping, and Discrimination* (pp. 199-227). New York, NY: Psychology Press.
- Jussim, L., Harber, K. D., Crawford, J. T., Cain, T. R., & Cohen, F. (2005). Social reality makes the social mind: Self-fulfilling prophecy, stereotypes, bias, and accuracy. *Interaction Studies*, 6(1), 85-102. doi: 10.1075/is.6.1.07jus
- Jussim, L. J., McCauley, C. R., & Lee, Y.-T. (1995). Why study stereotype accuracy and inaccuracy?. In Y. T. Lee, L. J. Jussim & C. R. McCauley (Eds.), *Stereotype Accuracy: Toward Appreciating Group Differences* (pp. 3-27). Washington, DC: American Psychological Association
- Karlins, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 13(1), 1-16. doi: 10.1037/h0027994
- Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, 28(3), 280-290. doi: 10.1037/h0074049
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98(2), 155-163. doi: 10.1037/0033-295X.98.2.155
- Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, 102(3), 390-402. doi: 10.1037/0033-2909.102.3.390
- Krueger, J. I. (2012). The (ir)rationality project in social psychology: A review and assessment. In J. Krueger (Ed.), *Social Judgement and Decision Making* (pp. 59-75). New York, NY: Psychology Press.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, 4:863. doi:10.3389/fpsyg.2013.00863
- Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality*, 42(4), 914-932. doi:10.1016/j.jrp.2007.12.003

- Letzring, T. D. (2010). The effects of judge-target gender and ethnicity similarity on the accuracy of personality judgments. *Social Psychology, 41*(1), 42-51. doi: 10.1027/1864-9335/a000007
- 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., 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*(1), 111-123. doi: 10.1037/0022-3514.91.1.111
- Lewis, K. L., Hodges, S. D., Laurent, S. M., Srivastava, S., & Biancarosa, G. (2012). Reading between the minds: The use of stereotypes in empathic accuracy. *Psychological Science, 23*(9), 1040-1046. doi: 10.1177/0956797612439719
- Li, J., & Chignell, M. (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human-Computer Studies, 68*(9), 589-602. doi: 10.1016/j.ijhcs.2010.04.001
- Löckenhoff, C.E., Chan, W., McCrae, R.R., De Fruyt, F., Jussim, L., De Bolle, M., ... Terracciano, A. (2014). Gender stereotypes of personality: Universal and accurate? *Journal of Cross-Cultural Psychology, 45*(5), 675-694. doi:10.1177/0022022113520075
- MacNell, L., Driscoll, A., & Hunt, A. N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, doi: 10.1007/s10755-014-9313-4
- Madon, S., Guyll, M., Hilbert, S. J., Kyriakatos, E., & Vogel, D. L. (2006). Stereotyping the stereotypic: When individuals match social stereotypes. *Journal of Applied Social Psychology, 36*, 178-205. doi: 10.1111/j.0021-9029.2006.00057.x
- Markey, P. M., & Wells, S. M. (2002). Interpersonal perception in internet chat rooms. *Journal of Research in Personality, 36*(2), 134-146. doi:10.1006/jrpe.2002.2340
- McCauley, C. R. (1995). Are stereotypes exaggerated? A sampling of racial, gender, academic, occupational, and political stereotypes. In Y. T. Lee, L. J. Jussim & C. R. McCauley (Eds.), *Stereotype Accuracy: Toward Appreciating Group Differences* (pp. 215-243). Washington, D.C.: American Psychological Association.
- McConahay, J. B. (1983). Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin, 9*(4), 551-558. doi:10.1177/0146167283094004

- McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-125). Orlando, FL: Academic Press.
- McCrae, R. R., Chan, W., Jussim, L., De Fruyt, F., Löckenhoff, C. E., De Bolle, M.,... Terracciano, A. (2013). The inaccuracy of national character stereotypes. *Journal of Research in Personality*, 47 (6), 831-842. doi: 10.1016/j.jrp.2013.08.006
- McKenna, K. Y. A., & Bargh, J. A. (2000). Plan 9 from cyberspace: The implications of the internet for personality and social psychology. *Personality and Social Psychology Review*, 4(1), 57-75. doi: 10.1207/S15327957PSPR0401\_6
- McLarney-Vesotski, A., Bernieri, F., & Rempala, D. (2011). An experimental examination of the “good judge”. *Journal of Research in Personality*, 45(4), 398-400. doi:10.1016/j.jrp.2011.04.005
- Mischel, W. (1968). *Personality & assessment*. New York, NY: John Wiley & Sons, Inc.
- Neville, H. A., Lilly, R. L., Duran, G., Lee, R. M., & Browne, L. (2000). Construction and initial validation of the color-blind racial attitudes scale (Cobras). *Journal of Counseling Psychology*, 47(1), 59-70. doi: 10.1037//0022-0167.47.1.59
- Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3), 211-236. doi:10.1080/01638530802073712
- Olson, M. A. (2009). Measures of prejudice. In T. Nelson (Ed.), *Handbook of Prejudice, Stereotyping, and Discrimination* (pp. 367-386). New York, NY: Psychology Press.
- Oltmanns, T. F., Friedman, J. N. W., Fiedler, E. R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behavior. *Journal of Research in Personality*, 38(3), 216-229. doi: 10.1016/S0092-6566(03)00066-7
- Ottati, V., & Lee, Y. -T. (1995). Accuracy: A neglected component of stereotype research. In Y. T. Lee, L. J. Jussim & C. R. McCauley (Eds.), *Stereotype Accuracy: Toward Appreciating Group Differences* (pp. 29-59). Washington, DC: American Psychological Association
- Otten, S., & Epstude, K. (2006). Overlapping mental representations of self, ingroup, and outgroup: Unraveling self-stereotyping and self-anchoring. *Personality and Social Psychology Bulletin*, 32(7), 957-969. doi: 10.1177/0146167206287254
- Pennebaker J.W., Francis M.E., Booth R.J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC 2001*. Mahwah, NJ: Erlbaum
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577. doi: 10.1146/annurev.psych.54.101601.145041

- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741-763. doi: 10.1037/0022-3514.67.4.741
- Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on twitter. *Journal of Research in Personality*, *46*(6), 710-718. doi: 10.1016/j.jrp.2012.08.008
- Radvansky, G. A., Copeland, D. E., & von Hippel, W. (2010). Stereotype activation, inhibition, and aging. *Journal of Experimental Social Psychology*, *46*(1), 51-60. doi:10.1016/j.jesp.2009.09.010
- Reid, S. A., & Hogg, M. A. (2005). Uncertainty reduction, self-enhancement, and ingroup identification. *Personality and Social Psychology Bulletin*, *31*(6), 804-817. doi: 10.1177/0146167204271708
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology*, *38*(1), 119-125. doi: 10.1002/1097-4679(198201)38:13.0.CO;2-I
- Rogers, K. H., & Biesanz, J. C. (2014). The accuracy and bias of interpersonal perceptions in intergroup interactions. *Social Psychological and Personality Science*, *5*(8), 918-926. doi:10.1177/1948550614537307
- Rule, N. O., & Ambady, N. (2011). Judgments of power from college yearbook photos and later career success. *Social Psychological and Personality Science*, *2*(2), 154-158. doi:10.1177/1948550610385473
- Rule, N. O., Tskhay, K. O., Freeman, J. B., & Ambady, N. (2014). On the interactive influence of facial appearance and explicit knowledge in social categorization. *European Journal of Social Psychology*, *44*(6), 529-535. doi:10.1002/ejsp.2043
- Rutgers University Office of Institutional Research and Academic Planning. (2016). *2015-2016 Rutgers University Fact Book :Enrollment* (pp. 5-25, Rep.).
- Schneider, D. J. (2004). *The psychology of stereotyping*. New York, NY: The Guilford Press.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515-530. doi: 10.1037/0022-3514.51.3.515
- Shrauger, J. S., & Osberg, T. M. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin*, *90*(2), 322-351. doi: 10.1177/01461672962212004
- Slepian, M. L., Bogart, K. R., & Ambady, N. (2014). Thin-slice judgments in the clinical context. *Annual Review of Clinical Psychology*, *10*, 131-153. doi: 10.1146/annurev-clinpsy-090413-123522

- Spence, J. T., Helmreich, R., & Stapp, J. (1973). A short version of the attitudes toward women scale (aws). *Bulletin of the Psychonomic Society*, 2(4), 219-220. doi: 10.3758/BF03329252
- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68(2), 199-214. doi: 10.1037/0022-3514.68.2.199
- Swim, J. K., & Hyers, L. L. (2009). Sexism. In T. Nelson (Ed.), *Handbook of Prejudice, Stereotyping, and Discrimination* (pp. 199-227). New York, NY: Psychology Press.
- Terracciano, A., Abdel-Khalek, A. M., Adam, N., Adamovova, L., Ahn, C., Ahn, H. N., . . . McCrae, R.R.(2005). National character does not reflect mean personality trait levels in 49 cultures. *Science*, 310, 96-100. doi:10.1126/science.1117199
- Tougas, F., Brown, R., Beaton, A. M., & Joly, S. (1995). Neosexism: Plus Ça change, plus c'est pareil. *Personality and Social Psychology Bulletin*, 21(8), 842-849. doi: 10.1177/0146167295218007
- Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and osn: A meta-analysis. *Journal of Research in Personality*, 49, 25-30. doi: 10.1016/j.jrp.2013.12.004
- University of California Berkeley Office of Planning & Analysis. (n.d.). *UC Berkeley Fall 2015 Enrollment Data*. Retrieved April 26, 2016, from University of California at Berkeley Office of Planning & Analysis website: <http://opa.berkeley.edu/uc-berkeley-fall-enrollment-data>
- Vazire, S. (2010). Who knows what about a person? the Self–Other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281–300. doi:10.1037/a0017908
- Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, 20(2), 104-108. doi:10.1177/0963721411402478
- Vazire, S., & Gosling, S. D. (2004). E-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, 87(1), 123-132. doi: 10.1037/0022-3514.87.1.123
- Williams, J.E., & Best, D.L. (1990) *Measuring sex stereotypes: A multinational study*. Newbury Park, CA: Sage
- Williams, M. J., & Mendelsohn, G. A. (2008). Gender clues and cues: Online interactions as windows into lay theories about men and women. *Basic and Applied Social Psychology*, 30(3), 278-294. doi: 10.1080/01973530802375136

Zaki, J., & Ochsner, K. (2011). Reintegrating the study of accuracy into social cognition research, psychological inquiry. *Psychological Inquiry*, 22(3), 159-182. doi: 10.1080/1047840X.2011.551743

Zou, L. X., & Dickter, C. L. (2013). Perceptions of racial confrontation: The role of color blindness and comment ambiguity. *Cultural Diversity and Ethnic Minority Psychology*, 19(1), 92-96. doi: 10.1037/a0031115