

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/279851095>

# Contrasting self-report and consensus ratings of intellectual humility and arrogance

Article in *Journal of Research in Personality* · July 2015

Impact Factor: 2 · DOI: 10.1016/j.jrp.2015.07.002

CITATION

1

READS

257

5 authors, including:



**Benjamin R. Meagher**  
Franklin and Marshall College

14 PUBLICATIONS 25 CITATIONS

SEE PROFILE



**Shawn J Latendresse**  
Baylor University

48 PUBLICATIONS 1,414 CITATIONS

SEE PROFILE



**Wade Rowatt**  
Baylor University

57 PUBLICATIONS 1,417 CITATIONS

SEE PROFILE



## Contrasting self-report and consensus ratings of intellectual humility and arrogance



Benjamin R. Meagher<sup>a,\*</sup>, Joseph C. Leman<sup>a</sup>, Joshua P. Bias<sup>b</sup>, Shawn J. Latendresse<sup>a</sup>, Wade C. Rowatt<sup>a,\*</sup>

<sup>a</sup> Baylor University, United States

<sup>b</sup> University of Mary Hardin-Baylor, United States

### ARTICLE INFO

#### Article history:

Received 12 January 2015

Revised 2 July 2015

Accepted 3 July 2015

Available online 6 July 2015

#### Keywords:

Intellectual humility

Intellectual arrogance

Person perception

Virtue

Positive psychology

### ABSTRACT

Despite a growing interest in intellectual humility (IH) and intellectual arrogance (IA), adequate measurement remains a challenging issue. This paper presents a pair of studies that compare two strategies: self-assessments and relational measures of group consensus. In Study 1, unacquainted participants provided round-robin judgments following a set of collaborative tasks. A social relations analysis revealed no consensus for either construct, making the relational measure untenable. However, a round-robin design following months of cooperative course work (Study 2) produced consensus for both constructs. Self-reported IH in both studies was positively associated with self-enhancement, despite the construct's definitional association with accurate self-appraisals, whereas relational IH was not. These studies reveal key ways in which personal and relational assessments can differ.

© 2015 Elsevier Inc. All rights reserved.

“Self-seeking, self-glory, that is not me. No. Many people say I embarrass them with my humility.” Archbishop Peter Akinola, as quoted in [Polgreen and Goodstein \(2006, December 25\)](#).

The study of human virtues has a long intellectual history, particularly in the fields of philosophy and theology. However, it has only been within the past few decades that the positive psychology movement led psychological researchers to begin to seriously consider virtues and their role in human life ([Peterson & Seligman, 2004](#)). The result has been a wealth of recent scholarship on a variety of relevant topics, such as gratitude and forgiveness ([Carlisle & Tsang, 2013](#)), love ([Fehr, 2013](#)), and self-control ([Baumeister & Vohs, 2012](#)). Despite this broad, growing interest in positive human attributes, humility, on the other hand, has been referred to as the “most overlooked and underappreciated virtue” ([Chancellor & Lyubomirsky, 2013, p. 819](#)), as it has yet to produce a comparably large body of work within empirical psychology.

This dearth has recently generated a great deal of consideration, motivated in large part by extended reflection on the critical importance of humility as a virtue specifically within the intellectual and academic domain ([Thrive Center for Human Development, 2014](#)). For example, the advancement of scientific knowledge seems to fully depend on practitioners possessing some degree of

intellectual humility. That is, researchers must be motivated to pursue the truth, wherever that may lead them, instead of being focused on status within the field, defending a “pet theory” regardless of its adequacy, or refusing to question one's own initial assumptions and positions in light of new, conflicting evidence ([Roberts & Wood, 2003](#)). Even among non-scientists, learning from others first requires an acknowledgment and admission of ignorance ([Hodges, Meagher, Norton, McBain, & Kimball, 2014](#)), so education itself is largely dependent upon these open expressions of intellectual humility.

In light of the wide-reaching influence humility has on critical aspects of human social functioning, empirical efforts to better evaluate and understand this construct are well overdue. In this paper, we begin with a brief description of the two primary challenges responsible for curtailing empirical research on the topic of humility generally: conceptual issues, in terms of defining humility, and measurement issues, regarding how one can accurately assess individual differences. Following this discussion, we consider the relevance of these theoretical and methodological issues for humility within the intellectual domain specifically.

### 1. Conceptual issues in the study of humility

The first stumbling block for an empirical approach to studying humility has been a basic conceptual question: What is humility? As is true for many terms in the psychological literature, conceptual definitions of humility often differ dramatically among lay

\* Corresponding authors at: Department of Psychology and Neuroscience, Baylor University, Waco, TX 76798, United States.

E-mail addresses: [benjamin\\_meagher@baylor.edu](mailto:benjamin_meagher@baylor.edu) (B.R. Meagher), [wade\\_rowatt@baylor.edu](mailto:wade_rowatt@baylor.edu) (W.C. Rowatt).

persons, theoreticians, and researchers. As noted by [Tangney \(2009\)](#), dictionary definitions typically describe humility as merely holding oneself in low regard, a trait entailing meekness, self-abasement, and low self-esteem. However, despite this fairly negative portrayal, lay theories of humility are actually quite positive, associating humility with good psychological adjustment and positive emotions ([Exline & Geyer, 2004](#)). This finding is consistent with the historical, English-language lexical research that led to the development of the Five Factor Model of personality structure ([Costa & McCrae, 1985](#); [McCrae & Costa, 1997](#)), which places humility-related items within the facet of modesty under the higher-order factor of agreeableness. Thus, this framework views humility as one component of having a prosocial and communal orientation toward others. More recently, Lee and Ashton (2004), developing their own model of personality structure, have argued that lexical studies across multiple languages indicate that humility (with honesty) represents its own unique factor independent of agreeableness, which is characterized by facets of sincerity, fairness, greed-avoidance, and modesty. Notably, distinguishing between agreeableness and Honesty–Humility is generally done by researchers interested in different forms of social morality: agreeableness predicting receptive forms (e.g., tolerance, forgiveness) and Honesty–Humility predicting more agentic forms (e.g., altruism, pro-sociality).

Outside of these lexical and lay theories of humility, a number of philosophers, theologians, and psychologists have sought to develop more nuanced definitions of the construct. Although several different conceptions have been proposed ([Chancellor & Lyubomirsky, 2013](#); [Davis, Worthington, & Hook, 2010](#); [Exline et al., 2004](#); [Tangney, 2009](#)), humility within these frameworks is consistently characterized as a multidimensional construct, most commonly including an accurate or moderate assessment of one's own abilities, being open to new ideas, having a low self-focus, and being able to acknowledge one's own mistakes. Notably, a large portion of this theoretical work has involved distinguishing humility from a number of closely related constructs. For example, although measures of modesty (e.g., self-reporting lower values on desirable traits than do knowledgeable others) have often been used as proxies for humility, [Tangney \(2009\)](#) argues that modesty is a narrower construct, involving a moderate estimate of one's abilities but lacking the openness and low self-focus characteristic of humility. [Exline et al. \(2004\)](#) make a slightly different distinction, suggesting that modesty is an exclusively social trait, entailing a particular type of self-presentation that may or may not be consistent with internal humility. In both cases, humility is conceptualized as a more expansive construct than modesty. Researchers have also argued that humility differs from certain conceptually-related negative attributes, such as high self-esteem or narcissism, which is characterized by feelings of grandiosity, an overestimation of self-importance, and a sense of entitlement. [Tangney \(2009\)](#) points out that although narcissistic people necessarily lack humility, it is less evident that people low on narcissism must also be high in terms of humility. For example, a person with low self-esteem will likely be low in narcissism, but also potentially low in humility, as they may be engaging in self-deprecation as a means of eliciting a positive social response and therefore still demonstrating a high self-focus.

The philosophical and theoretical literature on humility has dramatically outpaced the empirical work attempting to scientifically evaluate these theories. As evidenced in this short review, theoretical work has primarily sought to clarify many subtle definitional distinctions between humility and related constructs. However, more empirical testing is needed to evaluate the adequacy and robustness of these nuanced conceptual frameworks when observing how individuals actually describe themselves and others. As a result, the extent to which humility can be

empirically disentangled from numerous related traits, such as agreeableness, modesty, narcissism, and arrogance, is still an open question. The primary cause of this difficulty—concerns over measurement—will be described next.

## 2. Measurement issues in the study of humility

The second chief challenge for empirical work on the topic of humility has been the question of how to accurately measure the construct ([Davis et al., 2010](#)). As with most psychological traits, researchers have generally relied on self-report measures, such as the Honesty–Humility subscale of the HEXACO Personality Inventory ([Lee & Ashton, 2004](#)) or the Modesty–Humility subscale of the Values in Action Strengths Inventory ([Peterson & Seligman, 2004](#)). However, as the opening quote of this paper hints, it is unclear whether humility can be accurately self-reported. Would a humble person be likely to brag to a newspaper that he embarrasses others with his humility? The large-scale distribution of Rev. Akinola's quote by his many detractors would suggest that most people believe not.

This concern over self-report measurement stems from the very characteristics of the attribute itself. Because humility is conceived of as entailing an accurate or moderate view of oneself (i.e., not self-enhancing), as well as a low self-focus, it is a construct that is inherently linked to self-assessment. As a result, it is perhaps not surprising that many question the internal validity of a self-report measure. For example, people with low humility may self-enhance and report high levels, and people with high humility may express modesty and report lower levels ([Davis et al., 2010](#)). This challenge has led a number of researchers to develop and consider alternative measurement strategies ([Chancellor & Lyubomirsky, 2013](#)), such as implicit assessments ([Rowatt et al., 2006](#)). However, the most prominent alternative methodology employed in the past several years is the use of personality judgments from raters ([Davis et al., 2010, 2011, 2013](#); [Kruse, Chancellor, Ruberton, & Lyubomirsky, 2014](#)). Davis and colleagues ([Davis et al., 2010, 2011](#)) have framed this approach as a measurement of *relational humility*, defined as a social judgment, rather than an intrinsic individual attribute. They argue that humility is easier and more reliably assessed in others, as it avoids problems related to self-enhancement and socially desirable responding. Moreover, this perspective draws on [Vazire's \(2010\)](#) self-other knowledge asymmetry model, which proposes that ratings by others tend to be more accurate than self-assessments on traits that are highly evaluative, in that they are closely tied to motivational and ego-defensive processes. Humility, being a highly valenced construct, falls into this evaluative classification.

Nevertheless, two important challenges exist for the quantitative assessment of relational humility. First, Davis and colleagues' ([Davis et al., 2010, 2011, 2013](#)) relational model proposes that relational humility is best measured in terms of inter-judge agreement across a number of raters. However, these peer ratings will only be meaningful if there actually is consensus among raters ([Kenny, 1994](#)). Groups are most likely to reach consensus for trait judgments when perceivers witness the same or similar behaviors that reflect a particular trait ([Kenny, Albright, Malloy, & Kashy, 1994](#)), viz., actions that provide good information ([Funder, 1995](#)). A consensus assessment of humility may therefore be limited only to a very specific set of contexts or relationships that are capable of revealing this virtue. Several authors have argued that humility will be most evident behaviorally in situations where it is directly challenged, such as during interpersonal conflict, when receiving recognition or praise, when interacting with someone of a lower social status, or when describing past success ([Chancellor & Lyubomirsky, 2013](#); [Davis et al., 2010, 2011](#); [Kruse et al., 2014](#)).

For example, [Davis et al. \(2013\)](#) had participants in groups take part in a number of activities designed to reveal humility, such as describing their respective strengths and weaknesses as a leader. However, the effectiveness of this method in its ability to reveal humility has not been robustly tested across different situations or time periods. If judgments of a particular individual's humility are primarily idiosyncratic to each perceiver, a measure of inter-judge agreement will not be particularly meaningful.

Another potential issue for a relational model of humility is the fact that it is much more time intensive than self-report measures. It is therefore important to assess how this method relates to self-reports of the same construct to determine if it actually provides different, let alone better, information. Research on related, though typically negatively valenced, constructs have generally revealed modest positive correlations between self- and other-ratings. For example, several studies have tested self-other agreement for impressions of narcissism, finding small but statistically significant, positive correlations of  $r < .20$  ([Cooper, Balsis, & Oltmanns, 2012](#); [Lukowitsky & Pincus, 2013](#)). [Carlson, Vazire, and Oltmanns \(2011\)](#) found no correlation between self- and other-ratings of narcissistic traits between minimally acquainted individuals, but found significant values when rated by a close other ( $r = .15-.27$ ). Similarly, in a workplace context, [Johnson et al. \(2010\)](#) found modest correlations between employees' self-ratings of arrogance and those provided by peers ( $r = .19$ ) and supervisors ( $r = .13$ ). While clearly demonstrating a relationship, these results also indicate substantial differences between self- and other-assessments on self-evaluative constructs.

Despite the recent influx of alternative measures of humility, fewer studies have actually provided direct comparisons between self-ratings and other-ratings on humility specifically. Those that have reveal fairly inconsistent results. For example, [de Vries, Lee, and Ashton \(2008\)](#) found positive correlations between self and other-ratings on Honesty–Humility using the HEXACO Personality Inventory, but here too the magnitude of the correlation varied dramatically as a product of the dyad's relationship ( $r = .22$  for casual acquaintances to  $r = .60$  for romantic partners). [Rowatt et al. \(2006\)](#) also found a positive correlation between self-reported humility and ratings given by close others (e.g., friend, partner, or family member), although the strength of this association varied across different scale measures. In stark contrast, [Davis et al. \(2013\)](#) found a statistically significant negative self-other correlation ( $r = -.31$ ) among minimally acquainted participants, suggesting the possibility of a “modesty effect,” with truly humble people self-reporting lower humility. This variability in the relationship between self and other-ratings poses a number of significant questions. At the very least these results indicate that, in certain social contexts, self and other-ratings can be quite distinct and will likely predict very different outcomes.

### 3. Intellectual humility

In the research reviewed thus far, humility has been conceptualized as a general, global virtue. However, an additional conceptual debate is the extent to which the construct is better understood as context- or ability-specific, implying that there may be multiple forms of humility relevant in different psychological domains. That is, an individual may show high levels of humility in one facet of life (e.g., academic ability) but not necessarily in another (e.g., social relationships). The subdomain of humility that has garnered the most attention from psychologists ([Gregg & Mahadevan, 2014](#); [McElroy et al., 2014](#); [Samuelson et al., 2014](#)) and philosophers ([Roberts & Wood, 2003](#)) is intellectual humility (IH), an epistemic virtue tied to the realm of knowledge and ideas. Conceptually, IH has been proposed to entail an accurate or

moderate assessment of one's own intelligence, being receptive to the contributions and ideas of others, and being able to accept criticism about one's own ideas (e.g., [McElroy et al., 2014](#); [Samuelson et al., 2014](#)). High IH therefore involves having a low self-focus and little concern for status or ownership over particular ideas, instead caring more about the intrinsic value of knowledge and truth ([Roberts & Wood, 2003](#)).

As discussed at the beginning of this paper, IH has been proposed to represent a singularly important epistemic virtue, necessary for science, education, and learning generally ([Thrive Center for Human Development, 2014](#)). Whether one is observing a lab of scientists or a college study group, the extent to which each member is receptive to others' ideas, capable of accepting criticism, and grounded in their assessment of their own knowledge will presumably have important implications for the group's capacity to learn, improve, and cooperate successfully. However, even within this specifically intellectual domain, similar challenges regarding the conceptual definition and measurement of humility persist. For example, there is considerable debate in terms of how intellectual humility relates to or is distinct from intellectual arrogance (IA). One perspective argues that they merely represent the opposing ends of the same construct ([Gregg & Mahadevan, 2014](#)): “[IA is] the inclination to regard a belief as true because it is one's own. IH, conversely, would then be the inclination not to, or the disinclination to do so” (p. 8). In contrast, [Samuelson et al. \(2014\)](#) have argued that IH represents a “virtuous mean,” located between the vice of IA (claiming to know more than is merited) and the opposing vice of intellectual diffidence (claiming to know less than is merited). In support of this conceptual distinction, [Samuelson et al. \(2014\)](#) found that folk descriptions of prototypical IH and IA persons, though primarily inversely related, nevertheless demonstrated unique dimensions for epistemic motivations: “An IA person uses education in a prideful way to confer social status, while an IH person pursues education out of curiosity and love of learning” (p. 14).

An additional conceptual question is the extent to which IH is distinguishable from a more general, global measure of humility. Recent empirical work provides some initial evidence that IH represents a distinct subdomain capable of predicting unique variance above and beyond general humility on a number of related constructs ([Davis, Rice, McElroy, DeBlaree, & Choe, 2015](#)). Of course, the value of testing for unique statistical variance depends wholly on accurate measurement, and several recent attempts have been made to develop scale measures specifically for IH ([Krumrei-Mancuso & Rouse, in press](#); [McElroy et al., 2014](#)). For example, [McElroy et al. \(2014\)](#) developed a 16-item scale, consisting of items such as: “Is open to others' ideas,” “Is good at considering the limitations of their perspective,” and “Seeks out alternative viewpoints.” However, similar questions regarding the adequacy of self-ratings also apply to these assessments of IH, just as they do to general humility. To what extent do individuals accurately report their own IH? A relational measure of IH derived from peer assessments ([Davis et al., 2010, 2011](#)) may provide a valuable alternative method, but this approach has not yet been evaluated within this intellectual domain. Importantly, the adequacy of this measurement strategy will again depend on whether groups actually do reach consensus in their evaluations.

The goal of the present studies was to compare and contrast self-ratings of IH with the ratings of peers at both minimal acquaintance (Study 1) and following several months of interaction within an academic context (Study 2). Specifically, we sought to address three primary questions. First, under what circumstances are peer ratings useful in providing a reliable measure of IH? That is, under what conditions will significant levels of consensus be reached for this construct? In Study 1, unacquainted individuals took part in a group task designed to require creative thinking,

intellectual skills (i.e., math and verbal abilities), and consideration of one's own strengths and weaknesses—activities that should in theory be most likely to demonstrate high or low levels of intellectual humility. In Study 2, consensus for IH was assessed among groups of students enrolled in a college course using a team-based, classroom learning environment (Michaelsen, Knight, & Fink, 2002). This intensive amount of joint activity specifically within an academic context provides a test of whether an extended period of time working in groups on intellectual exercises is capable of providing sufficient information for groups to reach consensus about members. Although the relationship between self-ratings and informant ratings has been shown to differ by level of acquaintance (de Vries et al., 2008), it is unclear if the amount of inter-judge agreement also changes over time.

The second goal of these studies was to evaluate the extent to which ratings of IH within these contexts would be statistically distinct from judgments of other constructs known to be related, but proposed to be theoretically distinguishable. Specifically, we collected both self- and other-ratings of items assessing both IH and IA, in order to assess the extent to which they are distinguishable as traits, as proposed by Samuelson et al. (2014). In contrast, finding large overlap between these two constructs would lend greater support to theories placing these terms on either end of a single spectrum (e.g., Gregg & Mahadevan, 2014; Roberts & Wood, 2003). Also measured were a number of additional constructs and items associated with a variety of frameworks for understanding general humility, such as agreeableness, interpersonal dominance (i.e., meekness), and self-esteem.

The final goal of the present studies was to compare self-ratings and other-ratings of IH within these group contexts. This analysis involved assessing the magnitude and direction of the self-other correlation for each construct, as well as observing differences in terms of their respective relationships with two particular outcome variables. The first, self-enhancement, represents the extent to which participants rate themselves highly on a variety of positive attributes, controlling for how positively they generally view others and how positively other people view them (Kwan, John, Kenny, Bond, & Robins, 2004). Although humility is a more expansive construct than is modesty (Tangney, 2009), they nevertheless partially overlap in terms of being related to an unexaggerated and accurate view of oneself. Thus, a measure of intellectual humility should not be associated with a general bias to self-enhance on positive attributes. The second outcome, academic achievement, was assessed in Study 2 with each participant's individual course grade given by the instructor, as well as their peer evaluation grade provided by fellow group members. Previous work has revealed a positive correlation between students' course grades and an implicit measure of humility (Rowatt et al., 2006), as well as between judgments of team contribution and group ratings (Owens, Johnson, & Mitchell, 2013). The current study expands on these findings by providing a test of the association between academic performance and intellectual humility specifically, in terms of both explicit self-report and group consensus.

## 4. Study 1

### 4.1. Method

#### 4.1.1. Participants

One hundred thirty-five undergraduate students (92 women) participated in this experiment in exchange for course credit. Ages ranged from 18 to 23 years old ( $M = 19.04$ ), and the sample was 61% White/Caucasian, 13% Asian/Pacific Islander, 12% Black/African American, 11% Hispanic, and 3% other race or ethnicity.

#### 4.1.2. Measures and procedure

Unacquainted participants arrived to the study in groups ranging in size from three to five people. Each participant was given a name tag with a letter (A–E), and they were told that they would be completing several activities together as a group. For the first activity, participants were asked to reflect on their own strengths and weaknesses and then share briefly (less than 30 s) about one of each to the group. In the second task, the group completed a brainstorming activity: the experimenter described a scenario in which everyone in the world spontaneously developed two extra thumbs that appear on the other side of their pinky fingers (Bouchard & Hare, 1970). Participants were instructed to list as many changes as possible that might occur as a result, and each participant then shared their top two most creative changes with the rest of the group. Finally, for the last activity, participants were given five minutes to work together to complete three GRE problems as a group (two math questions and one verbal question). After five minutes, the experimenter reviewed the correct answers with the group, discussing what they had gotten right and wrong. Together, these tasks took approximately 20 min for each group to complete.

Immediately after the third activity, each participant then completed the experimental questionnaire, which required them to provide personality judgments of every member of the group, including themselves, on a set of 6-point Likert-type scales. The two constructs of primary interest were: (1) *intellectual humility*, which was measured with four items ( $\alpha = .77$ ; "Open to criticism of ideas," "Knows what he/she is not good at," "Can learn from others," and "Is intellectually humble"), and (2) *intellectual arrogance*, which was measured with three items ( $\alpha = .71$ ; "Arrogant," "Is closed-minded," and "Believes own ideas superior to other's ideas"). Consistent with previous theorizing (Gregg & Mahadevan, 2014; Samuelson et al., 2014), we treated IH and IA as distinct constructs, rather than a single bipolar attribute.

Several other additional constructs hypothesized to be relevant to group functioning were also collected. These included *interpersonal dominance* ( $\alpha = .75$ ; "Assertive" and "Tends to dominate discussion") and *competence* ( $\alpha = .73$ ; "Intelligent," "Good at public speaking," "Has strong math skills," and "Has strong verbal skills"), as well as the Big-5 personality traits, which were adapted from the Ten Item Personality Inventory (Gosling, Rentfrow, & Swann, 2003): *extraversion* ( $\alpha = .71$ ; "Extroverted, enthusiastic" and "Reserved, quiet"), *agreeableness* ( $\alpha = .56$ ; "Critical, quarrelsome," "Sympathetic, warm," and "Likeable"), *conscientiousness* ( $\alpha = .30$ ; "Dependable, self-disciplined" and "Disorganized, careless"), *neuroticism* ( $\alpha = .59$ ; "Anxious, easily upset," "Calm, emotionally stable," and "Depressed"), and *openness* ( $\alpha = .60$ ; "Open to new experiences, complex," "Conventional, uncreative," and "Creative"). Finally, several additional, exploratory measures were measured with single-items: "Likes to be center of attention," "Honest," "Funny," "Has high self-esteem," and "Is a good leader."

#### 4.1.3. Data preparation and analysis

The analysis was limited to participants who were unacquainted prior to the study session. As a result, groups of three people that contained a pair of acquainted individuals were dropped entirely from the analysis. This was true of four groups. For groups that contained four or more people, one of the acquainted participants was randomly selected to be removed from the dataset. This occurred in three groups. In total, this left a sample size of 120 participants divided across 32 groups.

Data was analyzed using the Social Relations Model (SRM; Kenny, 1994), a statistical method that allows for the partitioning of interpersonal judgments into three primary components: (1) the perceiver effect, or how a perceiver generally rates everyone, (2) the target effect, or how a target is generally rated by everyone,

and (3) the relationship effect, or the idiosyncratic way in which a particular perceiver uniquely sees a particular target. By partitioning judgments into these three sources of variance, one can assess of the relative influence of each these factors. For example, a large amount of variance attributable to the perceiver effect reflects the tendency for participants to generally rate all other members of their group in a similar way (i.e., assimilation). Alternatively, a large amount of variance attributable to the target effect would reflect consensus within the group: there is agreement about which people are seen as intellectually humble and which people are not. In this study, variance partitioning was done using the software program SOREMO (Kenny, 1998). Because SOREMO is unable to control for missing data, the mode response was imputed into all empty fields prior to conducting the variance partitioning procedure. Imputed data accounted for just 0.3% of all responses.

## 4.2. Results

### 4.2.1. Relative variance partitioning

The relative variance partitioning for all measured construct ratings are shown in Table 1. Significance tests for these values involve computing variances for each group in the sample, and then employing a one-way *t*-test to evaluate whether the means of the variances differ from zero. Groups did not achieve statistically significant levels of consensus for IH. In fact, the largest source of variance came from the perceiver, which was greater than the target variance,  $t(31) = 4.06$ ,  $p < .001$ , and marginally greater than the relationship variance,  $t(31) = 1.73$ ,  $p = .093$ . Similarly, no target variance was found for IA. Instead, significantly greater variance was attributable to the perceiver,  $t(31) = 2.87$ ,  $p = .007$ , and to the relationship,  $t(31) = 2.08$ ,  $p = .045$ . The magnitude of IA perceiver variance did not differ from relationship variance,  $t(31) = 0.589$ ,  $p = .560$ . Constructs that did achieve statistically significant levels of consensus included extraversion,  $t(31) = 4.33$ ,  $p < .001$ , interpersonal dominance,  $t(31) = 2.76$ ,  $p = .009$ , leadership,  $t(31) = 3.68$ ,  $p = .001$ , and being funny,  $t(31) = 3.70$ ,  $p = .001$ .

### 4.2.2. Correlates of self-ratings

We examined associations between self-reported IH, IA, and other measured traits. These correlations with the self-ratings on the other measured constructs are shown in Table 1, under self-self correlations (disattenuated for reliability). Notably, while

ratings of IH and IA were correlated, the magnitude of the correlation indicates that participants did differentiate between the constructs. Those rating themselves as intellectually humble also tended to rate themselves positively on most other attributes, including, most strongly, being agreeable, funny, honest, and emotionally stable. Interestingly, self-reporting as high on IA was negatively correlated with these items, and it was particularly associated lower levels of emotional stability and lack of conscientiousness. Both IH and IA were positively associated with interpersonal dominance. Because no consensus was reached in groups regarding IH or IA, equivalent correlations with other-ratings were not conducted.

## 4.3. Discussion

In this study, unacquainted groups of individuals engaged in a series of interpersonal tasks prior to providing their impressions of their fellow group members. These interpersonal judgments were then partitioned into distinct sources of variance in order to assess relative amounts of consensus, assimilation, or relational idiosyncrasy. Our results indicated that the social tasks used in this procedure were unable to reveal IH or IA, as essentially no consensus was reached by groups for these constructs. A very similar procedure was employed by Davis et al. (2013) to measure general humility relationally, but they did not report the relative magnitudes of perceiver and target variance found in their study. One substantial difference between the two studies was the length of time participants spent together: whereas the tasks of the current study took no more than half an hour to complete, participants in Davis et al. (2013) spent approximately three hours together. Kenny (2004) has noted that consensus can increase over this period of time following initial exposure, suggesting the need for a more concerted effort to evaluate what particular situations and tasks reliably elicit consensus on these judgments, as well as determining the dynamics of how consensus may change over time.

Because there was no consensus reached by groups regarding IH and IA, other-ratings could not be directly compared to self-ratings, as was initially planned. However, observing the correlations between the various self-ratings still raised several questions regarding validity. Conceptually, although IH and IA showed an expected negative correlation, the magnitude of this relationship suggests that participants did appear to distinguish between them when evaluating themselves. However,

**Table 1**  
Relative variance partitioning and self-rating correlations in Study 1.

	Relative variance partitioning				Self-Self correlations <sup>a</sup>	
	Perceiver	Target	Relationship	Error	IH	IA
Intellectual humility	.316**	.025	.154*	.505	–	–.352**
Intellectual arrogance	.177*	.000	.271*	.552	–.352**	–
Interpersonal dominance	.152*	.190*	.223*	.435	.241*	.339**
Competence	.180*	.086	.133*	.601	.337**	–.301*
Extraversion	.000	.271**	.306**	.423	.260*	.040
Agreeableness	.127*	.033	.160*	.679	.694**	–.443**
Conscientiousness	.000	.049	.176*	.775	.101	–.602**
Emotional stability	.234**	.000	.078*	.688	.449**	–.644**
Openness	.053	.089*	.186**	.672	.065	–.001
Leader	.279**	.203**	–	.518	.295**	–.055
Center of attention	.303*	.178	–	.520	.021	.361**
Honest	.541**	.000	–	.459	.474**	–.368**
Funny	.249**	.227**	–	.524	.501**	–.186*
High self-esteem	.300**	.066	–	.635	.325**	–.086

Note: Variance attributable to relationship is indistinguishable from error on single-item measures.

\*  $p < .05$ .

\*\*  $p < .01$ .

<sup>a</sup> Correlations are disattenuated based on scale reliability. Significance tests for self-self correlations are based on  $df = 87$ .

self-reported IH showed strong positive associations with a number of qualities one can regard as high in social value, such as competence, honest, and being funny. Its correlation with agreeableness was particularly substantial and consistent with a Five Factor Model framework (Costa & McCrae, 1985; McCrae & Costa, 1997), which understands humility as subsumed within agreeableness. In contrast, self-reported IA showed a relationship with generally low self-appraisals, with particularly substantial negative correlations with conscientiousness and emotional stability. Moreover, both self-reported intellectual humility and arrogance were found to be positively correlated with interpersonal dominance, a trait one would not expect to be related to being other-oriented (Davis et al., 2011; Tangney, 2009).

The direction of these associations seem to reflect the concerns raised by researchers regarding the validity of self-reporting in terms of humility: People with low humility may self-enhance broadly and report high levels of humility, whereas those with high humility may express modesty and report lower levels of humility (Davis et al., 2010). In this study, self-ratings appeared to reflect either a general positive evaluation, which included viewing oneself as high in IH, or a general negative evaluation, which including describing oneself as high in IA. The association between high self-reported IA and negative evaluations of oneself is particularly surprising in light of previous work on related, global constructs, which have found expected associations between self-report assessments on narcissism and excessively positive views of oneself (e.g., Carlson et al., 2011). However, it should be noted that the brief items used here to assess IA differ substantially from the items typically used to measure narcissism (e.g., NPI-16; Ames, Rose, & Anderson, 2006), in focusing on willingness to acknowledge others' ideas, rather than on more general feelings of importance and grandiosity.

Thus, the results of this initial study highlight a pair of challenges for measuring IH: (1) other-ratings at minimal acquaintance do not appear to be reliable, even following tasks designed to reveal these traits, and (2) self-ratings of IH were correlated with a number of unrelated, socially valued attributes, despite the fact that humility is theoretically defined in part as having less motivation to self-enhance (Davis et al., 2011). It is possible that the positive self-ratings associated with IH in this study were not due to a bias, but instead reflective of either accuracy (e.g., people high in IH genuinely are more agreeable, honest, and competent) or a tendency to see everyone more positively, including themselves. In Study 2, we evaluate self-enhancement by controlling for these alternative possibilities.

## 5. Study 2

The procedure employed in Study 1 failed to produce a significant amount of consensus within the tested groups of unacquainted individuals. Although previous research has found fairly stable levels of consensus over time for certain traits, at first acquaintance such agreement depends largely on having shared stereotypes regarding the given attribute (Kenny, 2004). It is possible that people simply do not share initial assumptions about high levels of IH. As a result, a longer period of interaction may be necessary to reveal this trait. In Study 2, we test this possibility by assessing personality judgments within groups of students who have worked together over several months. It was hypothesized that in this context – where participants have multiple shared experiences related to learning and social cooperation – groups would be more likely to form reliable, shared opinions about who is high or low in IH.

Secondly, because these assessments were collected within a classroom context among well-acquainted individuals, the

relationship between IH and a pair of important outcomes could also be assessed. First, we sought to evaluate the relationship between self-ratings and other-ratings of IH with a tendency to self-enhance on other positive attributes, employing the procedure developed by Kwan et al. (2004). This approach assesses the correlation between IH and a given construct, after controlling for the ratings of other, well-acquainted individuals, as well as the perceiver's tendency to rate everyone a particular way. If a positive association persists even after including these controls, it would reflect a positive self-bias. Secondly, the relationship between academic performance and IH could also be measured, both for self-ratings and other-ratings. Previous work has found a positive relationship between general humility and academic success (Rowatt et al., 2006), as well as positive team evaluation (Owens et al., 2013). In fact, Owens et al. (2013) found that perceived humility can compensate for low general intelligence when being evaluated by others, and groups whose leaders express humility show greater engagement and satisfaction. The present study offers a further test of this relationship by comparing self-report and others' ratings of intellectual humility and arrogance, both for individual academic performance and peer evaluations of performance.

### 5.1. Method

#### 5.1.1. Participants

All participants were undergraduate students enrolled in one of four upper level psychology courses (Health Psychology, Industrial-Organizational Psychology, Lifespan Development, or Theories of Personality) that utilized a team-based learning approach (TBL; Michaelsen et al., 2002). The overarching goal of this pedagogy is to encourage critical thinking, engagement, and the development of interpersonal and communication skills. TBL involves assigning students to groups at the beginning of the semester, with whom they complete a variety of tasks, both individual and collectively, over the duration of an academic semester. Initially, out-of-class reading assignments (e.g., textbook chapters, articles) are assessed using "Readiness Assurance Tests," which are first completed individually and then retaken collaboratively with fellow group members, who provide feedback on each member's work. The majority of time in-class is devoted to application exercises that require each team to form a joint decision about a specific problem based on their content knowledge, which they then report to and discuss with the rest of the class. Students earn credit for their individual performance and for the performance of their group, which are weighted by peer evaluations of their contributions to group productivity.

In the current study, participants were assigned to groups of four to six people. At the end of the semester, the experimenter explained the research to the students, asked for volunteer participation, and acquired informed consent. Of the 108 students (83 women) enrolled, 103 agreed to participate in the study divided across 23 team groups. Ages ranged from 18 to 30 years old ( $M = 21.08$ ), and the sample was 49% White/Caucasian, 24% Hispanic, 13% Black/African American, 9% Asian/Pacific Islander, and 5% other race or ethnicity.

#### 5.1.2. Measures and procedure

Each participant completed the same questionnaire employed in Study 1, which asked them to evaluate the other members of the group on 6-point Likert-type scales. Again, the constructs measured were: (1) *intellectual humility*,  $\alpha = .77$ , (2) *intellectual arrogance*,  $\alpha = .74$ , (3) *interpersonal dominance*,  $\alpha = .75$ , (4) *competence*,  $\alpha = .71$ , and (5) Big-5 personality traits: *extraversion*,  $\alpha = .75$ , *agreeableness*,  $\alpha = .61$ , *conscientiousness*,  $\alpha = .64$ , *neuroticism*,  $\alpha = .66$ , and *openness*,  $\alpha = .65$ . Finally, several additional constructs

were measured with a single-item: “Likes to be center of attention,” “Honest,” “Funny,” “Has high self-esteem,” and “Is a good leader.”

At the completion of the semester, students' final grades consisted of two scores: the total number of points they had earned for individual performance (e.g., tests, essays) and group performance (e.g., exercises). To promote personal accountability and decrease social loafing, peer evaluations from those in their group determines what percentage of the full group performance points students ultimately receive. This latter outcome involved asking students to distribute 100 points among the other members of the group, based on their contribution during group work over the semester. The instructor then scaled each student's average peer rating relative to a hypothetical equal distribution of the points (e.g., if 4 people were rated in the group, each person's average would be divided by 25), so that a 1 would mean they contributed exactly their fair share, a larger number means they contributed more than their fair share, and a number less than one means they contributed less than their fair share.

### 5.1.3. Data preparation and analysis

Variance partitioning was again done using SOREMO (Kenny, 1998). Empty fields, accounting for just 0.07% of all responses, were imputed with the mode response for that item.

## 5.2. Results

### 5.2.1. Relative variance partitioning

The relative variance partitioning for all measured construct ratings are shown in Table 2. Unlike in the previous study, groups showed statistically significant levels of consensus for both IH,  $t(22) = 3.29, p = .003$ , and for IA,  $t(22) = 3.38, p = .003$ . For IH, target variance was statistically smaller than perceiver variance,  $t(22) = 2.23, p = .036$ , but not smaller than the relationship variance,  $t(22) = 1.06, p = .302$ . For IA, target variance did not significantly differ in magnitude from perceiver variance,  $t(22) = 0.62, p = .541$ , but was marginally greater than the relationship variance,  $t(22) = 1.97, p = .061$ . Comparing these results to Study 1, consensus for IA was significantly higher in Study 2 with greater acquaintance,  $t(53) = 3.69, p < .001$ , but consensus for judgments of IH were not statistically significantly greater,  $t(53) = 1.47, p = .148$ .

### 5.2.2. Correlates of self and other-ratings

Correlations between self-reported IH and IA and the other measured constructs are shown in Table 3, under the “Self–Self” column. As in Study 1, ratings of IH and IA were strongly correlated

but not indistinguishable, with a disattenuated correlation of  $-.471$ . Also consistent with Study 1, those rating themselves as intellectually humble also tended to rate themselves positively on most traits, including competence, agreeableness, emotional stability, leadership, honesty, and being funny. This study also found a strong positive correlation between IH and openness, which was not observed in Study 1. High self-ratings on IA also predicted higher levels of reported competence, as well as higher dominance, higher desire to be the center of attention, and lower levels of emotional stability.

Because consensus was found in this study for both IH and IA, correlations could also be calculated for the target effects of IH and IA with the other measured judgments. These values indicate whether those rated highly by the group on one trait are also likely to be rated highly by the group on another trait. These values are shown in the “Other–Other” column of Table 3. The magnitude of the disattenuated correlations between IH and IA, as well as for Agreeableness, indicate that there was essentially no distinction between these constructs in terms of their target variance. That is, the consensus in other-ratings for IH was largely equivalent to their judgments of IA and agreeableness.

Finally, correlations between other-ratings for IH and IA with self-ratings on the other constructs are shown in the “Other–Self” column. These values indicate which self-ratings were associated with being viewed as high in IH or IA by the group. Self–other agreement for ratings of IH was positive but non-significant, disattenuated  $r = .214, p = .162$ , whereas self–other agreement for ratings of IA was statistically significant, disattenuated  $r = .394, p = .005$ . This indicates a modest level of agreement between an individual's appraisal of him or herself and the impressions held by others in the group. Groups tended to view people as intellectually humble who reported being low in dominance and high in agreeableness. Groups tended to view people as intellectually arrogant who reported being high in dominance, extraversion, and wanting to be the center of attention, but low in agreeableness and conscientiousness.

### 5.2.3. Relationship with self-enhancement

Although people who rated themselves as intellectually humble also tended to rate themselves more positively in general, it is possible that this relationship is the product of either: (a) their tendency to rate everyone positively (i.e., their perceiver effect), or (b) everyone viewing them positively (i.e., their target effect). In order to disentangle self-ratings from other-ratings, self-enhancement indices were calculated based on the procedure developed by Kwan et al. (2004). This index proposes that self-ratings are the additive result of how one generally sees others, how they are generally seen by others, and their unique view of themselves. Thus, unique self-ratings (i.e., their self-enhancement) can be computed by partialing out group-mean-deviated perceiver and target effects. The columns of Table 4 provide the partial correlations for these analyses, measuring the relationship between the self-rating for each construct with IH and IA self-ratings, controlling for the other SRM components of the outcome variable.<sup>1</sup> Reporting oneself as high in IH was generally associated with a more positive view, showing self-enhancement in terms of competence, agreeableness, openness, being a leader, and having high self-esteem. IA self-reports were associated with idiosyncratically higher ratings of oneself in competence.

<sup>1</sup> Self-enhancement can also be measured by creating a discrepancy score by subtracting the perceiver effect and target effect from self-ratings (e.g., Davis et al., 2013). However, discrepancy scores can potentially confound the effects of self-ratings and perceiver/target effects (e.g., humility can correlate with a discrepancy score only because it is strongly related to the self-rating, regardless of its relationship with the perceiver or target effects).

**Table 2**  
Relative variance partitioning in Study 2.

	Perceiver	Target	Relationship	Error
Intellectual humility	.238**	.105**	.158**	.499
Intellectual arrogance	.206**	.261**	.111**	.422
Interpersonal dominance	.040	.446**	.169**	.345
Competence	.135**	.125**	.162**	.578
Extraversion	.000	.438**	.226**	.336
Agreeableness	.068	.181**	.157**	.594
Conscientiousness	.159**	.254**	.124**	.463
Emotional stability	.250**	.016	.140**	.594
Openness	.141*	.072	.183**	.604
Leader	.149**	.387**	–	.464
Center of attention	.057	.268	–	.675
Honest	.382**	.099	–	.519
Funny	.281**	.212**	–	.508
High self-esteem	.245*	.175**	–	.581

Note: Variance attributable to relationship is indistinguishable from error on single-item measures.

\*  $p < .05$ .

\*\*  $p < .01$ .



**Table 3**  
Disattenuated correlations of IH and IA for self- and other-ratings in Study 2.

	Intellectual humility			Intellectual arrogance		
	Self–Self	Other–Other	Other–Self	Self–Self	Other–Other	Other–Self
Intellectual humility	–	–	.214	–.471**	–.979**	–.102
Intellectual arrogance	–.471**	–.979**	–.388*	–	–	.394**
Interpersonal dominance	.202	–.735**	–.446*	.470**	.865**	.467**
Competence	.599**	–.458	–.030	.410*	.600**	.173
Extraversion	.104	–.442**	–.344	.010	.589**	.378*
Agreeableness	.422**	.977**	.524**	–.302	–.896**	–.435**
Conscientiousness	.078	.412**	.282	–.210	–.314*	–.323*
Emotional stability	.332*	–	.328	–.332*	–	–.190
Openness	.611**	–	.043	–.093	–	.012
Leader	.446**	–.218	.115	.130	.359*	–.067
Center of attention	–.086	–	–.277	.450**	–	.285*
Honest	.253*	–	–.055	–.232	–	–.019
Funny	.310*	.074	.032	.178	.077	.113
High self-esteem	.264*	–.676**	.051	.218	.738**	–.028

Note: Self-rating correlations are disattenuated based on scale reliability. Other-ratings are disattenuated based on the reliability of the target variance estimate. Significance tests are based on the uncorrected correlations,  $df = 79$ .

\*  $p < .05$ .

\*\*  $p < .01$ .

**Table 4**  
Partial correlations of IH and IA on self-enhancement scores in Study 2.

	IH self-rating	IA self-rating	IH (w/IA items) other-rating
Intellectual humility	–	–.122	–
Intellectual arrogance	–.105	–	–
Interpersonal dominance <sup>b</sup>	.212	.212	.026
Competence	.422**	.245*	.019
Extraversion <sup>b</sup>	.156	–.079	–.008
Agreeableness <sup>b</sup>	.290**	–.162	.116
Conscientiousness	–.134	–.028	.003
Emotional stability <sup>a</sup>	.012	–.020	.090
Openness <sup>a</sup>	.311**	.127	–.015
Leader	.382**	.092	.067
Center of attention	–	–	–
Honest <sup>a</sup>	.084	–.135	–.021
Funny	.220*	.190	–.069
High self-esteem	.250*	.117	.187

Note: Self-enhancement values represent partial correlations between self-ratings of constructs with arrogance/intellectual humility after controlling for perceiver and target effects. Significance tests are based on  $df = 77$ .

<sup>a</sup> Enhancement measure partials out perceiver effect only.

<sup>b</sup> Enhancement measure partials out target effect only.

\*  $p < .05$ .

\*\*  $p < .01$ .

Moreover, the relationship between one's target effect on these constructs and the tendency to self-enhance on the various self-report items was also calculated. Because the target effects for IH and IA were indistinguishable, these scales were combined for this analysis to form a single IH value with IA items reverse coded. To assess self-enhancement, self-ratings for each construct were correlated with this IH target effect, controlling for the other SRM components of the variable. Unlike self-reported humility, group consensus was not found to be related to self-enhancement on any construct. Interestingly, only for self-esteem was there trend toward self-enhancement among those identified by the group as high in intellectual humility,  $p = .098$ .

#### 5.2.4. Relationship with academic performance and peer ratings

To evaluate how IH and IA related to academic performance, self and other-ratings of the two key constructs were correlated with both individual grades (i.e., evaluation on assignments

completed individually) and peer evaluation grades (i.e., evaluation given by group members for group assignments). Self-ratings of IH showed a non-significant, negative relationship with performance, with a disattenuated  $r = -.223$ ,  $p = .161$ , for individual grade and disattenuated  $r = -.148$ ,  $p = .265$ , for peer evaluation. Self-ratings of IA, on the other hand, had a statistically significant, positive association with individual grades, disattenuated  $r = .369$ ,  $p = .029$ , and a marginally significant association with peer evaluations, disattenuated  $r = .201$ ,  $p = .066$ . In contrast, other-ratings of IH showed a marginally significant, positive association with peer evaluations, disattenuated  $r = .380$ ,  $p = .055$ . However, there was no relationship between other-ratings of IH and individual grades, disattenuated  $r = .017$ ,  $p = .535$ .

#### 5.3. Discussion

This study expanded on Study 1 by evaluating judgments of IH among students who had been interacting with one another over several months within an academic setting. Unlike in the initial study, this social context proved sufficient to elicit consensus within groups regarding who among them were intellectually humble and arrogant. How did they accomplish this? Successful judgments of others' personality traits depend on having access to observable behaviors (Funder, 1995). It is for this reason that traits like extraversion, which are manifested in overt behavior (e.g., speaking frequently), tend to elicit the largest amounts of agreement. In this study, it appears that groups also used this interpersonal behavioral information to inform their humility judgments. Group judgments of IH were strongly correlated with self-reported dominance (negatively) and agreeableness (positively), and impressions of IA were related to greater extraversion, dominance, and desire for attention. Thus, in the present context of cooperative group coursework, IH was inferred largely from evidence of positive interaction and deference to others, whereas those rated highly in terms of IA were those who spoke frequently and dominated the direction of the group. This result provides some initial evidence for what types of behavior are used by observers to form impressions of this intellectual virtue.

Critically, relational judgments by the group showed a modest correlation with self-ratings on the two critical constructs. Although self-reports did have a positive directional relationship with group consensus for both IH and IA, only for IA was this

association statistically significant, indicating a clear distinction between the two methods. Notably, self-ratings for the tested items used in this study appeared to show greater conceptual nuance than peer judgments. That is, when evaluating themselves, correlations between judgments of IH, IA, and agreeableness, though substantial, were not indicative of conceptual overlap. In contrast, the correlations between the target effects of these constructs approached 1.0, particularly after correcting for attenuation. This indicates that at the group level, perceivers did not differentiate between these variables when assessing members. According to Vazire's (2010) Self-Other Congruency Model, self-ratings more accurately assess traits that are low in observability, whereas other-ratings are more accurate for judgments of traits that are high in evaluativeness. However, it is also possible that self- and other- trait judgments may differ not just in terms of their accuracy, but also in terms of their complexity and clarity. With less available information than people evaluating themselves, other-ratings may be more likely to exaggerate the overlap between constructs. As a result, self-ratings may be able to reflect the theoretical distinction between IH and IA argued by Samuelson et al. (2014), but observations by others may not. Alternatively, one could argue that IH and IA are in fact overlapping constructs (e.g., Gregg & Mahadevan, 2014), and the results of the other-ratings in this study merely demonstrate this fact.

The distinction between self-ratings and other-ratings on IH and IA was further highlighted by their respective relationships with the other measured constructs. Self-ratings of IH were positively associated with the majority of positively valenced attributes, including being competent, agreeable, open, a leader, and funny. Although a portion of this relationship may be attributable to the fact that these individuals tended to rate everyone more positively, these positive correlations remained even after controlling for participants' respective perceiver and target effects (Kwan et al., 2004). Thus, rating oneself as higher in IH actually predicted a tendency to self-enhance. Although it is theoretically possible that this could reflect accuracy (e.g., intellectually humble people really are more competent), the range of items showing a self-enhancement bias makes this proposal seem unlikely. For example, there is little reason to believe that intellectual humility predicts being truly funnier than your group actually appreciates. Moreover, both self-reported intellectual humility and arrogance were positively related to self-enhancing in terms of interpersonal dominance, a trait one would not expect to be related to being other-oriented (Davis et al., 2011; Tangney, 2009).

Relational ratings of IH, on the other hand, did not show a statistically significant correlation with self-enhancement on any construct. In other words, people identified by their group as high in IH tended to be those with moderate ratings of themselves. Moreover, correlations with academic performance differed dramatically between measures. Consistent with previous work employing implicit (Rowatt et al., 2006) and peer evaluations (Owens et al., 2013) of general humility, group consensus on IH was related to receiving a better grade in the class, but only for assignments evaluated by peers. In contrast, self-ratings of IH did not correlate with better grades, instead trending in a negative direction, whereas self-ratings of IA were related to superior grades. This result is quite surprising, as IA is theoretically defined as having an exaggerated view of one's intellectual ability and knowledge. Nevertheless, the observed relationship found here again appears to highlight a challenge inherent in assessing intellectual virtues and vices with self-assessments. People who did well in the course were those who generally rated themselves highly as thinking their ideas are superior to others'. In fact, their grades suggest that they were correct in this assessment. If one is genuinely more competent and has better ideas than others in the group, they may be justified in being close-minded, if their

ultimate goal is to secure a better grade for themselves. If, on the other hand, one has more prosocial motivations, such as helping others in the group get a better grade, sensitivity to a more complex array of values may be necessary for successful group functioning (Hodges et al., 2014). Thus, assessing group outcomes may provide a way to determine in what ways IA can hinder performance, rather than predict success.

## 6. General discussion

In this pair of studies, we sought to evaluate two predominant ways of assessing intellectual humility and intellectual arrogance within a group context: self-assessments and group consensus ratings. Both methods proved to have limitations. For self-ratings of IH, the primary issue remains validity: in both studies, high self-reported IH had a strong, positive association with self-enhancement on a number of socially valued attributes, despite the fact that definitions of humility entail having accurate or moderate views of one's knowledge and ability (Davis et al., 2010; McElroy et al., 2014; Tangney, 2009). Although it remains possible that people high in IH are also truly competent, better leaders, and funnier, the uniquely high self-appraisals across such a range of attributes makes this hypothesis seem rather unlikely. The alternative explanation, that high IH self-reports are driven in large part by socially desirable responding, is consistent with more general models of personality ratings that challenge the accuracy of self-reports on highly evaluative traits (Vazire, 2010).

Having said this, one limitation of the current studies is a reliance on a brief measure of IH that has not been empirically validated in prior work. Because the empirical study of intellectual humility is a new field of inquiry, there is yet to be an established self-report scale in use from which we could have drawn. McElroy et al. (2014) has only recently developed a 16-item informant scale of IH, which could conceivably be adapted to a self-report measure. A measure such as this, which also has multiple sub-scales, may provide a better test of how well a self-report measure of IH can predict expected outcomes. Having said this, there was in fact substantial overlap between the brief items used in the present studies and the larger scale developed by McElroy et al. (2014). The latter contains items that closely parallel those used here: e.g., "Is good at considering the limitations of their perspective," "Is open to competing ideas," and "Is open to others' ideas." Moreover, it is important to note that employing a longer, multi-item scale within a round-robin context can prove challenging for participants in larger groups, as this exponentially increases the time required to complete the study, thus making its application potentially limited. Nevertheless, future scale testing of measures such as these would benefit from making direct comparisons between how people evaluate themselves and how they are evaluated by others on these items.

In light of the concerns for self-report assessments, relational measures of humility have been proposed to provide a viable alternative method for assessing the construct (Davis et al., 2010, 2011, 2013). The current studies were an initial step toward testing this claim, specifically in the domain of IH. Supporting its adequacy, relational IH in Study 2 showed no correlation with self-enhancement and a positive relationship with peer evaluations of academic performance, consistent with previous work (Owens et al., 2013; Rowatt et al., 2006). However, a possible weakness of this approach was the high overlap between judgments of IH and other, conceptually related constructs (McCrae & Costa, 1997; Roberts & Wood, 2003; Samuelson et al., 2014). Although it is possible that IH and IA are genuinely two ends of a single construct, the high overlap also observed for evaluations of agreeableness do indicate that one limitation of the relational

measure may be reduced discriminant validity. Scales with a larger number of items may help to address this issue (McElroy et al., 2014), but the time required to complete long scales increases exponentially when conducting research with round-robin groups.

An additional concern for a relational measure of IH was the finding that the magnitude of consensus within these tested groups was fairly small. There was essentially no agreement at minimal acquaintance in Study 1, and consensus accounted for only about 11% of the variance after several months of collaborative group work. The viability of this method may therefore be limited, as a relational estimate can function usefully as an IH measurement only in the specific contexts where the group shows agreement (Kenny, 1994). As a result, a primary goal for future work must be to better theorize and assess how IH is revealed to others. Answering this question will involve two empirical directions. First, researchers should seek to determine the types of behaviors and other individual features people associate with IH. In the current work, the consensus found for IH and IA was related to self-ratings of extraversion, dominance, and a desire to be the center of attention. However, these impressions provide only a general sense of how IH manifests in behavior. A more fine-grained analysis or coding of interpersonal interactions and communication within groups, such as perspective taking (Epley, Keysar, Van Boven, & Gilovich, 2004; Moll & Kadipasaoglu, 2013), speech convergence (Pickering & Garrod, 2004), and behavioral synchrony (Lakin, 2013), may reveal the more subtle interpersonal information that reflects a focus on and openness to others' ideas. Secondly, researchers should propose and test the features of social situations expected to elicit and reveal IH to observers. Several authors have already proposed several situationally defined variables believed to elicit humility, such as during interpersonal conflict, interacting along across established social hierarchies, or when asked to reminisce about past achievements (Chancellor & Lyubomirsky, 2013; Davis et al., 2010, 2011; Kruse et al., 2014). However, these claims have not been empirically tested. Importantly, although Study 1 adopted certain features of the procedure used by Davis et al. (2013), it failed to reveal IH reliably to participating groups. What types of procedures, settings, and time periods are necessary to reliably measure IH interpersonally therefore remains an open question. This lack of research at the level of situations reflects a general dearth within social psychology and a failure to fully develop taxonomic understanding of the settings and interactions that characterize human life (Funder, 2009; Reis, 2008), despite their demonstrated importance in eliciting and revealing particular traits (Fleeson, 2007). Nevertheless, a handful of existing frameworks do exist to provide a starting point for this work; for example, in terms of dyadic motivations (Kelley et al., 2003), domains of sociality (Fiske, 1992), values that constrain behavior (Hodges, 2007; Hodges et al., 2014), or ecological units (Barker, 1968). Moreover, the recent development of relevant measurement tools, such as the Riverside Situational Q-Sort (Rauthmann et al., 2014; Sherman, Nave, & Funder, 2010), provide a means of systematically assessing and classifying situations for this purpose. These approaches can provide researchers with the language needed to begin hypothesizing the different scenarios in which IH is potentially necessary and likely to be detected reliably by perceivers.

## 7. Conclusion

The scientific study of IH is a field still in its infancy. As with the study of humility generally, a key initial focus of inquiry must therefore be to determine what is actually being measured when individuals report high or low levels of IH about themselves or others. The two studies described here indicate that

self-assessment and other-assessment appear to tap into fairly distinct qualities and outcomes. Self-ratings of IH were predictive of self-enhancement for a variety of traits in both of these studies, highlighting the fact that socially desirable reporting is a clear issue in the self-report of this construct. Relational IH ratings, on the other hand, were more consistent with expectations for the construct, showing no relationship with self-enhancement and a positive directional association with academic performance. Nevertheless, this approach is limited to social contexts where groups can actually reach consensus in their judgments, and the extent to which observers can disentangle impressions of IH with other, related constructs, such as IA and agreeableness, remains unclear. Thus, future work is needed to investigate just what behavioral or situational information specifies peer assessments of IH, in order to better understand how it is expressed and detected within group contexts.

## Acknowledgments

This research was supported by a generous Grant from The John Templeton Foundation (#29630). We thank Skylar Brannon, Pauline Minnaar, and Kristin Smart for their assistance in data collection in Study 1.

## References

- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, 40(4), 440–450. <http://dx.doi.org/10.1016/j.jrp.2005.03.002>.
- Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.
- Baumeister, R. F., & Vohs, K. D. (2012). Self-regulation and the executive function of the self. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 180–197). New York, NY, US: Guilford Press.
- Bouchard, T. J., Jr., & Hare, M. (1970). Size, performance, and potential in brainstorming groups. *Journal of Applied Psychology*, 54(1), 51–55.
- Carlisle, R. D., & Tsang, J.-A. (2013). The virtues: Gratitude and forgiveness. In K. I. Pargament, J. J. Exline, & J. W. Jones (Eds.), *APA handbook of psychology, religion, and spirituality* (vol. 1): Context, theory, and research (pp. 423–437). Washington, DC, US: American Psychological Association.
- Carlson, E. N., Vazire, S., & Oltmanns, T. F. (2011). You probably think this paper's about you: Narcissists' perceptions of their personality and reputation. *Journal of Personality and Social Psychology*, 101(1), 185–201. <http://dx.doi.org/10.1037/a0023781>.
- Chancellor, J., & Lyubomirsky, S. (2013). Humble beginnings: Current trends, state perspectives, and hallmarks of humility. *Social and Personality Psychology Compass*, 7(11), 819–833. <http://dx.doi.org/10.1111/spc3.12069>.
- Cooper, L. D., Balsis, S., & Oltmanns, T. F. (2012). Self- and informant-reported perspectives on symptoms of narcissistic personality disorder. *Personality Disorders: Theory, Research, and Treatment*, 3(2), 140–154. <http://dx.doi.org/10.1037/a0026576>.
- Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO personality inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Davis, D. E., Hook, J. N., Worthington, E. L., Jr., Van Tongeren, D. R., Gartner, A. L., Jennings, D. J., et al. (2011). Relational humility: Conceptualizing and measuring humility as a personality judgment. *Journal of Personality Assessment*, 93(3), 225–234. <http://dx.doi.org/10.1080/00223891.2011.558871>.
- Davis, D. E., Rice, K. G., McElroy, S. E., DeBlare, C., & Choe, E. (2015). Distinguishing intellectual humility and general humility. *The Journal of Positive Psychology*. <http://dx.doi.org/10.1080/17439760.2015.1048818>.
- Davis, D. E., Worthington, E. L., Jr., & Hook, J. N. (2010). Humility: Review of measurement strategies and conceptualization as personality judgment. *The Journal of Positive Psychology*, 5(4), 243–252. <http://dx.doi.org/10.1080/17439761003791672>.
- Davis, D. E., Worthington, E. L., Jr., Hook, J. N., Emmons, R. A., Hill, P. C., Bollinger, R. A., et al. (2013). Humility and the development and repair of social bonds: Two longitudinal studies. *Self and Identity*, 12(1), 58–77. <http://dx.doi.org/10.1080/15298868.2011.636509>.
- de Vries, R. E., Lee, K., & Ashton, M. C. (2008). The Dutch HEXACO Personality Inventory: Psychometric properties, self-other agreement, and relations with psychopathy among low and high acquaintanceship dyads. *Journal of Personality Assessment*, 90(2), 142–151. <http://dx.doi.org/10.1080/00223890701845195>.
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3), 327–339. <http://dx.doi.org/10.1037/0022-3514.87.3.327>.
- Exline, J. J., Campbell, W. K., Baumeister, R. F., Joiner, T., Krueger, J., & Kachorek, L. V. (2004). Humility and modesty. In C. Peterson & M. E. P. Seligman (Eds.),

- Character strengths and virtues: A handbook and classification* (pp. 461–475). New York, NY: Oxford University Press.
- Exline, J. J., & Geyer, A. L. (2004). Perceptions of humility: A preliminary study. *Self and Identity*, 3(2), 95–114. <http://dx.doi.org/10.1080/13576500342000077>.
- Fehr, B. (2013). The social psychology of love. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 201–233). New York, NY, US: Oxford University Press.
- Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99(4), 689–723.
- Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, 75(4), 825–862. <http://dx.doi.org/10.1111/j.1467-6494.2007.00458.x>.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652–670. <http://dx.doi.org/10.1037/0033-295x.102.4.652>.
- 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. <http://dx.doi.org/10.1016/j.jrp.2008.12.041>.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [http://dx.doi.org/10.1016/s0092-6566\(03\)00046-1](http://dx.doi.org/10.1016/s0092-6566(03)00046-1).
- Gregg, A. P., & Mahadevan, N. (2014). Intellectual arrogance and intellectual humility: An evolutionary-epistemological account. *Journal of Psychology and Theology*, 42(1), 7–18.
- Hodges, B. H. (2007). Values define fields: The intentional dynamics of driving, carrying, leading, negotiating, and conversing. *Ecological Psychology*, 19(2), 153–178. <http://dx.doi.org/10.1080/10407410701332080>.
- Hodges, B. H., Meagher, B. R., Norton, D., McBain, R., & Kimball, A. (2014). Speaking from ignorance: Not agreeing with others we believe are correct. *Journal of Personality and Social Psychology*, 106(2), 218–234. <http://dx.doi.org/10.1037/a0034662>.
- Johnson, R. E., Silverman, S. B., Shyamsunder, A., Swee, H.-Y., Rodopman, O. B., Cho, E., et al. (2010). Acting superior but actually inferior?: Correlates and consequences of workplace arrogance. *Human Performance*, 23(5), 403–427. <http://dx.doi.org/10.1080/08959285.2010.515279>.
- Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An atlas of interpersonal situations*. New York, NY, US: Cambridge University Press.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York, NY US: Guilford Press.
- Kenny, D. A. (1998). Social relations model program: Computer software and instructions downloaded from <<http://davidakenny.net/srm/srpm.htm>>.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8(3), 265–280. [http://dx.doi.org/10.1207/s15327957pspr0803\\_3](http://dx.doi.org/10.1207/s15327957pspr0803_3).
- Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the big five. *Psychological Bulletin*, 116(2), 245–258. <http://dx.doi.org/10.1037/0033-2909.116.2.245>.
- Krumrei-Mancuso, E. J., & Rouse, S. V. (in press). The development and validation of the comprehensive intellectual humility scale. *Journal of Personality Assessment*.
- Kruse, E., Chancellor, J., Ruberton, P. M., & Lyubomirsky, S. (2014). An upward spiral between gratitude and humility. *Social Psychological and Personality Science*, 5(7), 805–814. <http://dx.doi.org/10.1177/1948550614534700>.
- Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, 111(1), 94–110. <http://dx.doi.org/10.1037/0033-295x.111.1.94>.
- Lakin, J. L. (2013). Behavioral mimicry and interpersonal synchrony. In J. A. Hall & M. L. Knapp (Eds.), *Nonverbal communication* (pp. 539–575). Boston, MA, US: De Gruyter Mouton.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329–358. [http://dx.doi.org/10.1207/s15327906mbr3902\\_8](http://dx.doi.org/10.1207/s15327906mbr3902_8).
- Lukowitsky, M. R., & Pincus, A. L. (2013). Interpersonal perception of pathological narcissism: A social relations analysis. *Journal of Personality Assessment*, 95(3), 261–273. <http://dx.doi.org/10.1080/00223891.2013.765881>.
- McCrae, R. R., & Costa, P. T. Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509–516. <http://dx.doi.org/10.1037/0003-066X.52.5.509>.
- McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington, E. L. Jr., et al. (2014). Intellectual humility: Scale development and theoretical elaborations in the context of religious leadership. *Journal of Psychology and Theology*, 42(1), 19–30.
- Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2002). *Team-based learning: A transformative use of small groups*. Westport, CT: Praeger.
- Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience*, 7. <http://dx.doi.org/10.3389/fnhum.2013.00558>.
- Owens, B. P., Johnson, M. D., & Mitchell, T. R. (2013). Expressed humility in organizations: Implications for performance, teams, and leadership. *Organization Science*, 24(5), 1517–1538. <http://dx.doi.org/10.1287/orsc.1120.0795>.
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York, NY: Oxford University Press.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–226.
- Polgreen, L., & Goodstein, L. (2006). At axis of Episcopal split, an anti-gay Nigerian. *The New York Times*. Retrieved from <<http://www.nytimes.com>>.
- Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., et al. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, 107(4), 677–718. <http://dx.doi.org/10.1037/a0037250>.
- Reis, H. T. (2008). Reinvigorating the concept of situation in social psychology. *Personality and Social Psychology Review*, 12(4), 311–329. <http://dx.doi.org/10.1177/1088868308321721>.
- Roberts, C. R., & Wood, W. J. (2003). Humility and epistemic goods. In M. DePaul & L. Zagzebski (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology* (pp. 257–279). New York, NY: Oxford University Press.
- Rowatt, W. C., Powers, C., Targhetta, V., Comer, J., Kennedy, S., & Labouff, J. (2006). Development and initial validation of an implicit measure of humility relative to arrogance. *The Journal of Positive Psychology*, 1(4), 198–211. <http://dx.doi.org/10.1080/17439760600885671>.
- Samuelson, P. L., Jarvinen, M. J., Paulus, T. B., Church, I. M., Hardy, S. A., & Barrett, J. L. (2014). Implicit theories of intellectual virtues and vices: A focus on intellectual humility. *The Journal of Positive Psychology*, 1–18. <http://dx.doi.org/10.1080/17439760.2014.967802>.
- Sherman, R. A., Nave, C. S., & Funder, D. C. (2010). Situational similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology*, 99(2), 330–343. <http://dx.doi.org/10.1037/a0019796>.
- Tangney, J. P. (2009). Humility. In S. J. Lopez & C. R. Snyder (Eds.), *Oxford handbook of positive psychology* (2nd ed., pp. 483–490). New York, NY, US: Oxford University Press.
- Thrive Center for Human Development (2014). Intellectual Humility Grants. Retrieved from <<http://thrivecenter.org/intellectualhumility/intellectual-humilityawardees/>>.
- 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. <http://dx.doi.org/10.1037/a0017908>.