Effect of Construal Communality on the Congruence between Self-Report and Personality Impressions

John A. Johnson
Pennsylvania State University
College Place
DuBois, Pennsylvania 15801 USA

Abstract

Kenny’s rater agreement model conceptualizes “shared meaning” as the extent to which raters agree that an act implicates a particular level of a particular trait. This study applies the shared meaning concept to item responses on the Hogan Personality Inventory (HPI). Seventy-four research participants first completed the HPI under normal instructions. Next they rated the level of extraversion, agreeableness, conscientiousness, emotional stability, and intellectual openness implied by the endorsement an HPI item. These ratings enabled the assessment of “construal communality,” defined as the degree to which any judge’s ratings of trait implication matched the mean ratings from all judges. It was hypothesized that persons with high construal communality are more likely to provide endorsement patterns that correlate with personality ratings made by knowledgeable acquaintances. Construct communality was operationalized in several ways: with a measure of Euclidian distance, a city-block measure of distance, a multiplicative model of proximity, and with Q-correlations between a judge's rating profile and the modal rating profile. Only for the extraversion dimension were correlations between HPI scores and acquaintance ratings significantly higher for participants possessing higher levels of construal communality as measured by the Euclidian and city block distance models. Trends for the other Big Five factors were not statistically significant and results for multiplicative and Q-correlation models of communality were inconsistent. I discuss the relevance of the notion that item responses are performative speech acts for designing items that will maximize test validity.

This is a virtually unpolished first draft of a paper I will be presenting at the symposium, Personality Judgments: Theoretical and Applied Issues, chaired by Peter Borkenau and Frank M. Spinath, to be held during the 11th European Conference on Personality, in Jena, Germany, July 25, 2002. Changes to this paper may occur before that date, and only portions of the paper will actually be presented, so quote at your own risk. Comments and suggestions would be greatly appreciated.
Effect of Construal Communality on the Congruence between Self-Report and Personality Impressions

Attribution theorist and personality trait critic, Ned Jones (1979), warned us that it is a "rocky road from act to disposition" (p. 107). Nonetheless, to form an impression of personality, we must travel this road. We cannot instantaneously perceive someone's personality in its entirety. To form an impression of personality we must integrate our perceptions of individual acts into dispositional Gestalts. The road from act to personality disposition is indeed rocky, but if we can use models such as David Funder's (1995) Realistic Accuracy Model (RAM) to understand the process, perhaps we upgrade the rocky road into a smoother boulevard.

My own research on the road from acts to dispositions has focused narrowly on one particular type of act: the speech act. A speech act is the production of spoken or written words whose primary purpose is to modify the present situation, including the behavior and state of mind of those to whom the words are directed (Allan, 1994). Speech act theorists stress that people only rarely use words merely to convey information, to describe or depict the way things are (Austin, 1962). In fact, purely propositional communication may exist only as an ideal in scientific discourse (van Oort, 1997). When ordinary people perform speech acts, their intent is almost always pragmatic influence on their social environment. To understand how dispositions can be deduced appropriately from speech acts, I think we need to consider how an actor intends to influence the audience.

To my knowledge, the only personality researcher who has recognized the relevance of speech act theory for traveling the road from acts to dispositions is Jerry
Wiggins. (Although Boele de Raad, 1999, relied on a speech act classification of verbs to construct a taxonomy of interpersonal trait verbs.) In his classic essay, "In Defense of Traits," Wiggins (1974/1997) suggests that we identify trait-relevant acts by following what speech act theorist John Searle (1969) calls a *constitutive rule*. This rule follows the form:

"X counts as Y in context C."

Constitutive rules apply to acts generally, not just speech acts. Wiggins gives the following example of a constitutive rule for a non-speech act: "An action (pushing) that is likely to harm or injure another (X) counts as aggressive (Y) in the context of rules for classifying the consequences of social actions (C)" (p. 101). People, in turn, are described as "aggressive" if, under certain circumstances, they have behaved in a manner likely to harm or injure another.

Note that the starting point of dispositional attribution is the description of an act in terms of its social consequences. Two independent observers of an act must apply the same constitutive rule to reach the same conclusions about the personality trait implied by the act. David Kenny (1994) refers to common constitutive rules as "shared meaning systems," and suggested that lack of shared meaning contributes to disagreement among judges of personality. One can measure the amount of shared versus unshared meaning within a panel of judges, as Peter Borkenau (1990) did, by telling research participants that a fictitious person A has performed act X and then asking the participants to what extent that Person A has Trait Y." Borkenau found that correlations between judges' prototypicality ratings ranged from .31 to .70 (average correlation = .53), leading him to
conclude that "unshared meaning systems are a major source of any lack of consensus that is found for ratings of personality" (Borkenau, 2001, p. 106).

The usual solution for reducing the impact of "unshared meaning systems" is to aggregate ratings across judges and to assume that the average rating represents our best approximation to reality (Block, 1961/1978; Hofstee, 1994). Furthermore, if we like we may eliminate any individual judge's ratings from the aggregate ratings if those ratings do not correlate sufficiently with the consensus of the remaining judges. This procedure is identical to dropping items with low item-total correlations during scale development.

The present research extends, in several novel ways, Borkenau's research on measuring shared meaning through prototypicality ratings. Borkenau asked research participants to imagine that a fictitious Person A has performed Act X and then had participants rate the extent to which Person A had Trait Y. In my study, I asked research participants to imagine that a fictitious person had just endorsed an item on a personality inventory, such as "It is easy for me to talk to strangers." A basic premise of this procedure is that agreeing or disagreeing with items on standard, self-report personality inventories is a type of speech act (Johnson, 1997a). In other words, I am assuming that there is no important difference between a person saying "I agree that it is easy for me to talk to strangers" and a person marking "agree" on a Likert scale in response to the personality item, "It is easy for me to talk to strangers."

If the speech act conception of item responses is correct, then respondents are not merely disclosing information, but are intending to create an impression on an audience (Johnson, 1981). The impression that respondents intend to make by an item response is called the *illocutionary force* of the speech act. The *actual* impression formed by
observers of the item response is called the *perlocutionary* force of the speech act. To measure the perlocutionary forces of item responses in my study, I had participants rate the extent to which item endorsements indicated a low or high level of each broad trait in the Five Factor Model: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellectual Openness.

In the past, researchers have usually measured the perlocutionary force of item responses to assess the subtlety of items (e.g., Duff, 1965; Holden & Jackson, 1979). The overwhelming consensus of this research is that scales constructed from items with clear, univocal, perlocutionary force predict acquaintance ratings of personality better than scales containing items without clear perlocutionary force (Johnson, 1993). The present research focuses, however, not on the differences in perlocutionary force of items, but, rather, differences in respondents' abilities to perceive the perlocutionary force of item responses. Gergen, Hepburn, and Fisher (1986) have already demonstrated that, with encouragement, sophisticated language users can plausibly interpret item responses as reflecting an indeterminate number of psychological traits. The question investigated here is whether persons who perceive the perlocutionary force of item responses as other people do provide more predictive responses to items than people with idiosyncratic perceptions of the perlocutionary force of items.

The present study uses the term *construal communality* to describe the degree to which a judge’s ratings of perlocutionary force (trait implication) corresponds to the mean ratings from all other judges. Construal communality might be seen as a specific manifestation of the tendency to interpret human behavior as others do. This tendency has been called *commonality* by George Kelly (1955), *communality* by Harrison Gough.
Construal Communality (1968) and perceptual conformance by Ted Sarbin (Sarbin and Hardyck, 1955). The research here examines the hypothesis that construal communality moderates the ability of self-report personality scores to predict relevant acquaintance ratings.

**Method**

*Participants*

Participants were 74 undergraduates (24 male, 50 female) enrolled in an introductory level psychology course taught by the author, plus two acquaintances chosen by each student to rate their personalities. The students received extra credit equivalent to 5% of their marks for the course for participating.

*Measures*

Participants completed the original, 310-item version of the Hogan Personality Inventory (HPI; Hogan, 1986) under standard instructions. The original HPI was the first personality inventory designed explicitly to assess the five major psycho-lexical factors described by Norman (1963), factors that are now called the "Big Five" (Goldberg, 1993). During scale construction, Hogan decided to develop two scales to assess Factor I (which Norman called Extroversion or Surgency): Sociability and Ambition. Subsequent factor analyses (Hogan, 1986; Johnson, 1994) show that Ambition items load in complex ways on Factors III, IV, and V as well as Factor I; therefore, the more univocal Sociability scale was chosen in the present study to represent Factor I. The remaining HPI scales correspond to the rest of the Big Five factors in straightforward fashion: Factor II, Likeability; Factor III, Prudence; Factor IV, Adjustment; and Factor V, Intellectance.
All samples were rated by acquaintances who knew them well with the Bipolar Adjective Rating Scales (BARS; Johnson, 1997a; Johnson & Ostendorf, 1993; Johnson, Germer, Efran, & Overton, 1988). The BARS was designed explicitly to gather acquaintance ratings of the same constructs measured by the original HPI (Hogan & Johnson, 1981). The BARS contains 49 items in a 7-step Likert format, anchored by two trait terms (e.g., imaginative vs. down-to-earth). Item ratings are summed to yield scores corresponding the constructs measured by the HPI, although unique scale names were devised to distinguish the BARS from the HPI: I. Sociality; II. Likeableness; III. Discipline; IV. Poise; and V. Mentality. Details on the construction, reliability, and validity of the BARS can be found in the aforementioned references. For the present study it is sufficient to know that the reliabilities of the scales are acceptable in all samples (Cronbach alphas between .70 and .90). Acquaintances returned their ratings in sealed envelopes through the students or mailed them directly to me.

After the students completed the HPI and distributed copies of the BARS to two acquaintances, they engaged in an item rating task similar to the procedure described by McCrae, Costa, and Piedmont (1993) (exact procedure with all instructions available upon request from author). Participants were given descriptions of the Big Five factors and told that they were being asked to rate, on a five-point scale, what endorsement of a personality item would imply about each of the factors. They were asked to write a -2 next to an item if they perceived the item to strongly imply the low end of the personality factor, a -1 if the item somewhat implied the low end, a 0 if the item was irrelevant, a +1 if the item somewhat implied the high end, and a +2 if the item strongly implied the high end of the factor. They were given several examples of personality items whose
endorsement might imply different levels of one or more of the factors. Here is one of the examples:

This is how you would rate an item that might seem to indicate strongly the positive pole of Conscientiousness and somewhat the negative pole of Agreeableness:

If I were a manager I would have no trouble being strict to make sure that those I supervised did their jobs correctly.

| If I were a manager I would have no trouble being strict to make sure that those I supervised did their jobs correctly. | EXT AGR CON STB INT |
|---|---|---|---|---|---|
| 0 | -1 | 2 | 0 | 0 |

**Analyses**

*Constructing a profile of perlocutionary force.* The inter-rater reliability among the 74 judges over all 1550 trait-implication ratings, estimated by Cronbach's alpha, was .94. Corrected judge-total correlations ranged from -.02 to .69 (average $r = .43$ using Fisher's $r$-$z$ transform). Judges' ratings of the trait implications of the HPI items were averaged to form a composite profile of perlocutionary force of the HPI items. A second composite profile was formed by using ratings only from judges whose rating profile correlated $r = .30$ or better with the ratings from the remaining judges. The ratings from this reduced set of judges ($n=54$) was only slightly more reliable (alpha=.95; average judge-total $r = .52$), and the mean ratings correlated substantially ($r = .98$) with the mean ratings from the full set. One test of the main hypothesis using the full set of ratings and the same test using the more reliable set of ratings produced the same results. A decision was made to base all further analyses on ratings from the reduced, more reliable set of
ratings. To ascertain that the measured perlocutionary force of the items reflected the HPI scoring keys, mean profile ratings for items on the five HPI scales were computed and compared.

Methods for measuring construal communality. Four methods for measuring construal communality—two described by Chaplin and Panter (1993) and two additional methods described by Borkenau (1990)—were used. The first model suggested by Chaplin and Panter for measuring similarity between two profiles of scores is Cronbach and Gleser's (1953) $D^2$ statistic. $D^2$ is actually a measure of distance or dissimilarity between profiles and takes into account the dissimilarity in two profiles' intra-profile mean and variance as well as the dissimilarity of the profiles' direction of scores across items. Cronbach and Gleser refer to these three aspects of the profile as elevation, scatter, and shape. $D^2$ is computed by summing the squared differences between corresponding scores from the two profiles.

The square root of $D^2$, $D$, is often used instead of $D^2$ because its distribution tends to be less skewed and less heavily influenced by larger differences between profiles. By the Pythagorean theorem, $D$ represents the actual geometric difference between raters in $n$-item-dimensional space. $D$ as a measure of Euclidean distance was used by Borkenau (1990). Because $D$ does not follow a normal distribution, sometimes it is normalized by dividing it by the square root of the number of items in the profile. This procedure was followed by Chaplin and Panter (1993).

For the present study, it makes no difference whether $D^2$, $D$, or normalized $D$ is used, because these alternate measures of construal communality correlate .99 with each other. Thus, when participants were divided into two groups at the median, the groups
were the same across the three measures and produced identical results in subsequent analyses. For convenience, $D$ will be used from this point on to refer to the Euclidian distance between a judge's rating profile and the composite profile of perlocutionary force.

An alternate distance measure, described by Borkenau (1990), is the city-block measure. Rather than summing the squared differences between scores as in the $D^2$ measure, the city-block model sums the absolute value of the differences between scores. Although the metric for the city block model avoids the squaring in the $D^2$ model, the concept is basically the same and similar results should be expected.

Borkenau describes what he calls a multiplicative model of proximity to assess the similarity between sets of scores, and this model differs significantly from the Euclidian and city-block models. Whereas the previous model gives equal weight to (dis)similarity between all scores, the multiplicative model weights more heavily the similarity of scores that differ from zero. Multiplicative proximity, which I will abbreviate here by the letters MP, is computed by summing the products of scores. Borkenau (1990) explains, in fine detail, the different implications of the Euclidian and multiplicative measures of (dis)similarity. I will provide one example to indicate how the two methods can produce different results in the context of the present study.

Consider hypothetical item cited earlier as an example for the judges in their rating task, "If I were a manager I would have no trouble being strict to make sure that those I supervised did their jobs correctly." Let us compare the ratings of two judges, A and B, to the mean ratings from the entire panel of judges:

<table>
<thead>
<tr>
<th></th>
<th>EXT</th>
<th>AGR</th>
<th>CON</th>
<th>STB</th>
<th>INT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean profile across 54 judges</td>
<td>0</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Ratings from Judge A</td>
<td>1</td>
<td>-2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ratings from Judge B</td>
<td>0</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

By the Euclidian measure, Judge A's distance from the mean profile is 

\[ ((-1-0)^2 + (-2-1)^2 + (2-2)^2 + (1-0)^2 + (0-0)^2)^{\frac{1}{2}} = \sqrt{3.5} = 1.732. \]

Judge B's distance is given by 

\[ ((0-0)^2 + (-1-1)^2 + (2-2)^2 + (0-0)^2 + (0-0)^2)^{\frac{1}{2}} = 0. \]

That is, Judge B's ratings exactly match the mean profile for a zero distance, much closer in Euclidian space than Judge A's ratings.

However, the multiplicative model gives the following proximity score for Judge A: 

\[ (1 \cdot 0) + (-2 \cdot -1) + (2 \cdot 2) + (1 \cdot 0) + (0 \cdot 0) = 6. \]

Judge B's ratings show the following MP score: 

\[ (0 \cdot 0) + (-1 \cdot -1) + (2 \cdot 2) + (0 \cdot 0) + (0 \cdot 0) = 5. \]

By MP model, Judge A is not penalized for rating the implications of the item as somewhat extraverted and emotionally stable, because these ratings are nullified when multiplied by the zero values in the mean profile. And Judge A is actually rewarded for rating the item even more disagreeable than the mean profile and therefore shows a profile that is more similar to the mean than Judge B's profile, according to the MP measure.

Borkenau (1990) explains the conditions under which the multiplicative model might fit data better than the Euclidian model. In the present study, however, I hypothesized that the Euclidian measure of profile similarity would moderate self/acquaintance correlations better than the multiplicative measure of personality for the following reason. A respondent who perceives several personality dimensions implied by an item is more likely to endorse the item for the "wrong" (non-keyed) reason, but the multiplicative model will downplay that mistake by weighting it with a near-zero value.
The Euclidian measure, on the other hand, requires a person to correctly identify what the item does not imply as well as what it implies to maximize the similarity score. Because HPI items are scored on only one Big Five scale, respondents who perceive only the keyed Big Five trait implication of an item are more likely to respond appropriately. Therefore, the Euclidian similarity measure is predicted to moderate self/acquaintance correlations more powerfully than the multiplicative measure.

Stephenson (1952) coined the term Q-correlation to refer to the product-moment correlation between two persons' scores. In the present context, the Q-correlation between a judge's item ratings and the profile of mean ratings looks like an intuitively appropriate measure of construal communality and was therefore employed in the study. Cronbach and Gleser (1953) demonstrated, however, that the Q-correlation is simply an inverse function of $D^2$ computed on standardized scores, i.e., scores with the differences in elevation and scatter removed. Chaplin and Panter (1993) present the expression for the Q-correlation between two profiles as $1-(D^2/2)$. Cronbach and Gleser note that the Q-correlation has been used successfully in research (see Block, 1961/1978, for examples), but caution that removing differences in elevation and scatter is generally unadvisable. In the present context, raters such as the hypothetical Judge A who perceives several trait implications rather than one major implication will produce profiles that are more elevated and flatter than the mean profile. Standardizing will reduce spurious "somewhat implied" ratings closer to zero and accentuate the more extreme "strongly implied" ratings. For this reason, the Q-correlation measure of construal communality could be expected to yield results similar to those obtained with the multiplicative measure of construal communality.
Moderator analyses. To explore whether judges' levels of construal communality might be different for each of the Big Five factors, separate Euclidian $D$s were computed for extraversion, agreeableness, conscientiousness, emotional stability, and intellectual openness ratings. The average inter-correlation among the $D$s was .82. This indicates that a judge whose ratings of extraversion implication for the 310 HPI items tended to match the group's mean extraversion ratings also tended provide ratings of agreeableness, conscientiousness, emotional stability, and intellectual openness that matched the group means. Therefore, rather than using separate $D$s for each factor, an overall $D$ was computed across all 1550 ratings. One could suggest as a reliability estimate for this overall $D$ the average inter-correlation of .82 for the 310-item measures, corrected by the Spearman-Brown formula to .96.

Separate construal communality scores for each of the Big Five factors were also computed for the city-block ($CB$) model, the multiplicative proximity ($MP$) model, and the Q-correlation ($QC$) measure of profile similarity. In each case, strong inter-correlations among the 310-item measures suggested using an overall $CB$, $MP$, and $QC$ index based on all 1550 ratings. Average inter-correlations and corrected reliability estimates for each construal communality index are as follows: $CB$ (.84, .96), $MP$ (.74, .93), and $QC$ (.97, .97). The four overall measures of construal communality ($D$, $CB$, $MP$, $QC$) were inter-correlated to examine their similarities and differences.

Participants were divided into two groups at the median value for each construal communality measure. Within the below-median and above-median groups, correlations were computed between the following HPI scales and their corresponding BARS scales: HPI Sociability with BARS Sociality; HPI Likeability with BARS Likeableness; HPI
Prudence with BARS Discipline; HPI Adjustment with BARS Poise, and HPI Intellectance with BARS Mentality. Next, the more exacting moderated regression approach (Tellegen, Kamp, & Watson, 1982) was used to assess the moderating effect of construal communality across the range of predictor scores. Finally, exploratory correlations were conducted between the four construal communality measures and the HPI and BARS to see whether construal communality could be defined in terms of traditional personality variables.

Results

As a manipulation check, profiles of judges' ratings were computed for items on the HPI Sociability, Agreeableness, Prudence, Adjustment, and Intellectance scales. Table 1 summarizes the ratings. This summary confirms that in every case judges' ratings are in line with the standard HPI keying. For example, judges rated HPI Sociability items higher in Extraversion than the remaining four dimensions of the Big Five, and judges gave higher Extraversion ratings to HPI Sociability items than to the other four HPI scales.

Intercorrelations among the four measures of construal communality are shown in Table 2. As expected, the Euclidian D and city-block (CB) measures were highly related, as were the multiplicative proximity (MP) and Q-correlation (QC) measures. Also, not surprisingly, the QC measure of profile similarity, because it is an inverse function of standardized distance, showed moderate negative relationships to the D and CB measures of profile dissimilarity.

The findings relevant to the main hypothesis of the study are presented in Table 3. Although there appears to be a trend for higher correlations between HPI scores and
acquaintance ratings for individuals higher in construal communality when defined by the Euclidian $D$ (average $r = .51$ vs. .30) and city-block distance measures (average $r = .39$ vs. .32), the difference between high-communality and low-communality individuals is statistically significant only for Extraversion. The difference between r-to-z transforms divided by the standard error of the difference is 2.588 for the Extraversion correlations moderated by $D$ ($p < .01$, two-tailed) and 2.18 for the Extraversion correlations moderated by $CB$ ($p < .05$, two-tailed). Differences for the multiplicative and Q-correlation moderators do not consistently favor high- or low-communality individuals, and none of the differences are statistically significant. Series of moderated multiple regressions (Tellegen, Kamp, & Watson, 1982) confirm that only the Extraversion correlation is significantly moderated the distance measures, and only marginally ($p$ for $D$ beta weight $= .053$ and $p$ for $CB$ beta weight $= .086$).

Exploratory correlations between the construal communality measures and all of the HPI and BARS scales failed to turn up strong personality correlates of construal communality. Only two HPI Adjustment facets (No Somatic Anxiety and Calmness) and one HPI Likeability facet (Trusting) correlated significantly (-.25 to -.34 range, $ps < .05$) with $D$ and $CB$. The distance measures did not correlate significantly with any of the BARS full scales, but did correlate significantly with the individual items "progressive" and "introverted." With the large number of correlations (over 50 HPI facet scales and 49 BARS items) the personality correlations could be easily attributed to chance.

Discussion

*Construal communality* was defined in this study as the tendency to judge the personality implications of inventory item endorsements—a type of speech act—
similarly to the typical judgments of others. I suggested that for individuals to respond appropriately (validly) to personality inventory items, they must understand at some level what kind of personality impression an item endorsement makes on other people in general. Hence, it was predicted that correlations between inventory scores and peer acquaintance ratings of the same personality dimensions would be strongest for individuals with high levels of construal communality.

Support for the main hypothesis was weak, reaching statistical significance only for the Extraversion dimension. Trends in the predicted direction were achieved for the Agreeableness, Emotional Stability, and Intellectual Openness dimensions, but did not reach statistical significance. It is difficult to judge whether the null findings are accurate or reflect the low power of the relatively small sample size. For example, with the current sample size of \( N = 37 \) in each group, the power of the test is only .34 for judging that a correlation of .48 is significantly larger than .30 at the .05 level (one of the actual comparisons in the study). More definitive conclusions will require more data from larger samples. Fortunately, the study indicates that small subsets of item ratings are sufficiently reliable for measuring construal communality, so future research participants will not be faced with the onerous task of making 1550 HPI item ratings. Ratings from the present study could be used to select a sample of HPI items with a range of perlocutionary equivocality for future research.

Given the small sample size in this study, I do not want to draw strong conclusions about the empirical side of this research, but I do want to conclude by discussing some issues involving the application of speech act theory to personality assessment. First, the hypothesis that construal communality affects the validity of
personality test scores assumes that test takers are in fact responding to the items in order to create personality impressions rather than merely describing their actual thoughts, feelings, and behavior. In Austin's (1962) original treatise on speech acts, he distinguished verbal utterances he called *constatives*, which describe an objective state of affairs, from what he called *performatives*, which create a state of affairs by their very utterance. An archetypal example of a performative is the act of naming something, as in "I hereby christen this vessel H.M.S. Pinafore." In contrast, a constative verbalizes a proposition such as, "The name of the ship in that opera is H.M.S. Pinafore." One might reasonably ask whether endorsements of personality items such as "I am fun-loving person" are constatives—mere description—or performatives—performances designed to create impressions on others.

Most personality psychologists have treated personality item endorsements as if they are (or should be) constatives rather than performatives (Johnson, 1981, 1997a). They have assumed that item responses should veridically describe a person's actual thoughts, feelings, and behavior. It is up to the test scorer and interpreter to make valid inferences about personality from these veridical reports, just as a medical doctor diagnoses diseases from patients' simple self-reports of symptoms. Constatively oriented psychologists would no more expect a person to answer items with an eye toward creating a personality impression than a doctor expects a patient to self-diagnose his or her medical condition. Psychologists coming from a constative view have assumed that responses guided primarily by attempts to create a certain impression (i.e., performative responses) constitute a "response style" that would interfere with valid assessment. The lone dissenters to this constative view of personality item response have been those who
focus on the empirical correlates rather than the content of personality items (Buchwald, 1961; Meehl, 1945) and a few researchers who assume that personality item responses are in fact performatives (Hogan, 1996; Johnson, 1997a; Taylor, Carithers, & Coyne, 1976). So how do we decide whether item responses are constatives or performatives?

If Austin's (1962) analysis of language is correct, it would be a mistake to think that personality item responses are either constatives or performatives, because Austin eventually concludes that constatives represent a special case of the performative rather than a different type of language.¹ I made the same point about item responses (Johnson, 1981) when I suggested that "the self-presentation [performative] view incorporates the self-disclosure [constative] perspective, but goes beyond it" (p. 761). Austin concludes that all language is, at bottom, performative in the sense that people use words primarily to have an effect on other people (van Oort, 1997). Uttering constatives is simply one particular technique for influencing other people.

To illustrate how an apparent constative can have a performative function, consider the statement, "Personality psychologists understand human nature better than social psychologists." On the face of it, this looks like a constative: a proposition that is objectively true (or false). However, if I announce that I agree with this statement, my audience cannot help to form impressions of my personality. Social psychologists might see me as arrogant, disagreeable, and stupid, while my colleagues in personality might see me as insightful, sensible, and wise.

Now, if this item for some reason appeared on a personality inventory, how should I respond to it to create a valid impression of my personality? What matters is neither what my actual, private opinions are about personality and social psychologists
nor whether my response corresponds to my actual, private opinions. What matters is whether the impression created by my response corresponds to my "actual" personality (defined as the way most people see me—Hofstee, 1994). If most people regard me as very arrogant, disagreeable, and stupid, and social psychologists are scoring this personality test, then I should strongly agree with the statement.²

Psychologists holding the constative view of item responses probably find the above performative view of item responses disturbing for two reasons. First, the performative view might be construed as a form of social contractionism that denies the reality of a stable personality. After all, my analysis above suggests that my real, actual opinion of personality and social psychologists is irrelevant to producing a valid item response. What counts is how my performance impresses my audience, and different audiences will be impressed differently by the same response. If every item response is "just" a performance, an act contrived to manipulate the thoughts of an audience, how can item responses ever reveal someone's real personality?

I think this first worry is unfounded because it mistakenly equates all performing with manipulative lying (Chriss, 1995; Schlenker & Weigold, 1990) and therefore incorrectly assumes that there is no standard for judging whether a performance communicates a person's authentic personality. But we do have a standard—specifically, whether the personality impression created by the performance matches the person's established reputation. Established reputation—a social fact—is easily measured by aggregating ratings from judges who are well-acquainted with the person (Block, 1961/1978; Hofstee, 1994). One can, like preeminent speech act theorist John Searle
Construal Communality     20

(1992, 1995), assume that socially constructed facts represent "overlays" upon a real world that exists independently of what humans think about it.

But what are we to make of the fact that sometimes different audiences form different impressions of personality from the same performance? If my response to the hypothetical personality item leads social psychologists to regard me as arrogant, disagreeable, and stupid but leads personality psychologists to think of me as insightful, sensible, and wise, how can one determine how sensible and wise I am really? The performative view of item responses would answer, "With a better item," that is, with an item that creates a more consistent personality impression. It is true that performances of all types (not just speech acts) can create different impressions on different audiences, but personality psychologists generally do not study that phenomenon. Personality is usually defined in terms of consistencies (Johnson, 1997b), including consistent impressions people make on other people in general. Whereas a constantive view of personality items (e.g., Wolfe, 1993) recommends items that maximize accurate self-description, the performative view advocates items that maximize the consistency of perlocutionary force.

If we can accept the idea that personality test-takers might consider the perlocutionary force of their item responses, this does not, of course, mean they will necessarily choose responses that are consistent with their established reputations. Respondents who are aware of the perlocutionary force of item responses can choose to mislead us. To the extent that this happens, this would lower validity coefficients between personality scale scores and acquaintance ratings for persons high in construal communality. In the present study, validity coefficients were not consistently lower for
persons high in construal communality, so widespread misrepresentation must not be occurring.

The second question that I expect from traditional, constative-oriented psychologists is whether the performative view in fact fits the phenomenology of individuals responding to personality items. To what extent does the typical test-taker actually think, "Will the person scoring this test form an accurate impression of my personality if I answer the item this way?" versus "will my answer accurately describe my actual thoughts, feelings or behavior?"

Speech act theorists might answer this question in several different ways. First, Johnson (1990) and Searle (1995) have both suggested that often-practiced, rule-governed behaviors eventually become so second nature that people can follow a rule without being aware that they are following the rule. Therefore, people responding to items might be following constitutive rules that imply personality implications without being consciously aware that they are doing this (Mills & Hogan, 1978). Long ago I suggested (Johnson, 1981, Footnote 1) that further research is needed to clarify the extent to which item responses are conscious and reflective versus unconscious and reflexive, and how conscious and unconscious factors affect the validity of item responses. Despite some excellent work in this area (Paulhus, 1984; Paulhus & Reid, 1991) since that time, I believe that much additional research on this topic is still needed.

Another way to look at this issue is to remember that, if Austin's analysis is correct, constatives are a type of performative, so respondents do not necessarily have to choose between thinking about their item response in terms of either self-descriptive accuracy or perlocution. They could think simultaneously about both. Furthermore, with
the types of items used in actual, existing personality inventories, conflicts between the literal truth of an item response and how the response is scored are rare. Test-takers can (and often do) create valid personality impressions by taking a constative approach and endorsing true propositions about their thoughts, feelings, and behavior.

Finally, the performative view does not really predict that everyone will consider the perlocutionary force of their item response options. Many—even most—respondents might in fact focus on the self-descriptive accuracy of their responses rather than how their responses will be judged. Instructions to personality inventories do, after all, encourage respondents to think constatively. The question is, how risky is it to expect a personality test-taker to act like the apocryphal scientist who disinterestedly describes things as they really are? I think that the potential for problems depends on the types of items we use. If we use simple items whose perlocutionary force is tied to one clear, literalistic interpretation of an item, we can expect respondents to act like constative scientists. On the other hand, items containing any sort of non-literalistic, figurative language need to be approached performatively. An example will illustrate this.

Consider an item, “I am usually on time for meetings.” The constative view says that a conscientious person who is indeed usually on time for meetings will compare that behavior to the item’s content and then endorse it as literally accurate. The performative view says that a conscientious person understands that endorsing the item will be interpreted as a sign of conscientiousness and therefore endorses it because it communicates his/her conscientiousness. On the other hand, consider the item, “I am never late for meetings.” A purely literal-minded, conscientious person who was once late for a meeting could not, from the constative viewpoint, endorse the item because it is
not literally true. However, a conscientious person with good perspective-taking skills and high construal communality might well endorse the item because he/she understands that the response will be interpreted as a sign of conscientiousness.

In other words, cognitive style (literal versus figurative thinking) could interact with different types of items (Johnson, 1993). I have demonstrated that, under certain conditions, endorsing true propositions can lead to invalid personality impressions and stating false propositions can lead to valid personality impressions (Johnson & Horner, 1990). I therefore think we will come out much further ahead if we accept that language is naturally performative and then design personality items that encourage people to perform in ways that faithfully reflect how most people view their personalities.
References


Author Note

John A. Johnson, Department of Psychology, Pennsylvania State University, DuBois.

This research was first reported in an invited symposium, Personality Judgments: Theoretical and Applied Issues, chaired by Peter Borkenau and Frank M. Spinath, held during the 11th European Conference on Personality, in Jena, Germany, July 25, 2002. I thank Lew Goldberg for his interpretive input and Susan Butler for her assistance in coding the data into electronic format.

Correspondence concerning this article should be addressed to John A. Johnson, Penn State DuBois, College Place, DuBois, Pennsylvania 15801. E-mail: j5j@psu.edu.
Footnotes

1Searle (1979) eventually recognized constatives as performatives by renaming them *assertives* or *representatives*.

2In reality, because personality scores are normally derived from responses to a set of items rather than a single item, it doesn't matter whether my response to a single item is literally true or whether it is scored in the direction of my actual personality. What matters is whether my responses to all of the items on the scale produce a score that corresponds to my actual level on a personality dimension.

3If item responses are performances, then an individual’s social skills and abilities (e.g., perspective-taking) will determine how well the performances are executed. Thus, we have a conception of personality item responses with a resemblance to an "abilities conception" of personality (Wallace, 1966), although here we are talking about items responses as skilled performances rather than about skilled performances (e.g., Riemann, 1996).
Table 1

*Personality Impressions of Items on Five Hogan Personality Inventory Scales*

<table>
<thead>
<tr>
<th>HPI Scale</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>S</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociability</td>
<td>Mean</td>
<td>0.91</td>
<td>0.21</td>
<td>-0.06</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.49</td>
<td>0.29</td>
<td>0.19</td>
<td>0.25</td>
</tr>
<tr>
<td>Likeability</td>
<td>Mean</td>
<td>0.67</td>
<td>0.93</td>
<td>0.29</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.57</td>
<td>0.42</td>
<td>0.21</td>
<td>0.33</td>
</tr>
<tr>
<td>Prudence</td>
<td>Mean</td>
<td>-0.15</td>
<td>0.16</td>
<td>0.58</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.43</td>
<td>0.29</td>
<td>0.40</td>
<td>0.32</td>
</tr>
<tr>
<td>Adjustment</td>
<td>Mean</td>
<td>0.23</td>
<td>0.08</td>
<td>0.12</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.40</td>
<td>0.21</td>
<td>0.32</td>
<td>0.44</td>
</tr>
<tr>
<td>Intellectance</td>
<td>Mean</td>
<td>0.06</td>
<td>0.09</td>
<td>0.33</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.18</td>
<td>0.11</td>
<td>0.24</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Note. $N=74$. E, A, C, S, and I refer to judges' ratings of an item's implications for Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellectual Openness, respectively. Boldface values indicate highest ratings out of five dimensions for each scale on the Hogan Personality Inventory.
Table 2

*Intercorrelations Among Four Measures of Construal Communality*

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>CB</th>
<th>MP</th>
<th>QC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidian Distance (D)</td>
<td>(.96)$^a$</td>
<td>.99***</td>
<td>.09</td>
<td>-.50***</td>
</tr>
<tr>
<td>City Block Distance (CB)</td>
<td>(.96)$^a$</td>
<td>.12</td>
<td>-.45***</td>
<td></td>
</tr>
<tr>
<td>Multiplicative Proximity (MP)</td>
<td>(.93)$^3$</td>
<td>.77***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q-Correlation (QC)</td>
<td>(.97)$^a$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $N=74$.  

$^a$Numbers in parentheses represent reliability estimates.  

$^{***}p < .001$ (one-tailed)
### Table 3

**HPI-BARS Correlations Moderated by Four Measures of Construal Communality**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Criterion</th>
<th>Euclid D</th>
<th>City Block</th>
<th>Multiplicative</th>
<th>QCorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Full</td>
<td>Far</td>
<td>Close</td>
<td>Far</td>
</tr>
<tr>
<td>HPI</td>
<td>BARS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociability</td>
<td>Sociality</td>
<td></td>
<td>.32**</td>
<td>.05</td>
<td>.59***</td>
</tr>
<tr>
<td>Likeability</td>
<td>Likableness</td>
<td>.41***</td>
<td>.39*</td>
<td>.42*</td>
<td>.35*</td>
</tr>
<tr>
<td>Prudence</td>
<td>Discipline</td>
<td>.49***</td>
<td>.50***</td>
<td>.50***</td>
<td>.49**</td>
</tr>
<tr>
<td>Adjustment</td>
<td>Poise</td>
<td>.39***</td>
<td>.30</td>
<td>.48**</td>
<td>.36*</td>
</tr>
<tr>
<td>Intellectance</td>
<td>Mentality</td>
<td>.36***</td>
<td>.24</td>
<td>.53***</td>
<td>.25</td>
</tr>
</tbody>
</table>

*Note. N=74 for the full sample and N=37 for each far and close subsamples.*

*p < .05 **p < .01 ***p < .001 (all two-tailed).