Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

The Social Accuracy Model of Interpersonal Perception: Assessing Individual Differences in Perceptive and Expressive Accuracy

Jeremy C. Biesanz

To cite this article: Jeremy C. Biesanz (2010) The Social Accuracy Model of Interpersonal Perception: Assessing Individual Differences in Perceptive and Expressive Accuracy, Multivariate Behavioral Research, 45:5, 853-885, DOI: 10.1080/00273171.2010.519262

To link to this article: http://dx.doi.org/10.1080/00273171.2010.519262

Published online: 15 Nov 2010.

Multivariate Behavioral Research, 45:853–885, 2010

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273171.2010.519262

Psychology Press
Taylor & Francis Group

2007 Cattell Award Address Paper

The Social Accuracy Model of Interpersonal Perception: Assessing Individual Differences in Perceptive and Expressive Accuracy

Jeremy C. Biesanz
University of British Columbia

The social accuracy model of interpersonal perception (SAM) is a componential model that estimates perceiver and target effects of different components of accuracy across traits simultaneously. For instance, Jane may be generally accurate in her perceptions of others and thus high in *perceptive accuracy*—the extent to which a particular perceiver's impressions are more or less accurate than other perceivers on average across different targets. Just as well, Jake may be accurately perceived by others and thus high in *expressive accuracy*—the extent to which a particular target is accurately perceived on average across different perceivers. Perceptive and expressive accuracy can be further decomposed into their constituent components of normative and distinctive accuracy. Thus the SAM represents an integration of Cronbach's componential approach with Kenny's (1994) social relations model. The SAM is illustrated using both a half-block as well as a round-robin design. Key findings include reliable individual differences in several specific aspects of interpersonal perceptions.

The impressions and judgments that people make of others' personalities help to understand and explain past behavior and predict future behavior (e.g., Funder,

Correspondence concerning this article should be addressed to Jeremy C. Biesanz, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4. E-mail: jbiesanz@psych.ubc.ca

1991, 1995; Olson, Roese, & Zanna, 1996). There is substantial literature documenting consensus among personality judgments, agreement with self-reports, and the predictive validity of such judgments (e.g., Funder & West, 1993; Kenrick & Funder, 1988). For example, observers reach consensus on some personality traits even with very minimal information such as several seconds or minutes of observation of a target person and these consensual judgments predict the target person's future behavior (e.g., Albright, Kenny, & Malloy, 1988; Ambady & Rosenthal, 1993; Borkenau & Liebler, 1992; Norman & Goldberg, 1966). The overall accumulation of evidence demonstrates that, on average, people are able to accurately judge the personalities of others. They are able to form accurate impressions of personality where accuracy is defined following Funder (1995, 1999) as agreement between knowledgeable informants, the observer and the target, or between the observer or target and behavioral measures. However, are some people better judges of personality than others (the good judge)? Are some individuals more accurately perceived than others (the good or transparent target)? These are classic questions raised in the 1940s and the 1950s that have never been satisfactorily answered. Given the centrality of personality judgments in everyday life as well as psychological research, understanding when and for whom such judgments are more reliable, informative, and accurate remains a critical question.

Assessing accuracy in interpersonal perception has historically presented methodological challenges. In a series of critiques of earlier research, Cronbach (1955; Gage & Cronbach, 1955) introduced a methodological framework to help strengthen inferences of judgmental accuracy. Cronbach (1955) critiqued single global measures of accuracy such the Euclidean distance between judgments and a criterion and proposed instead to examine different components of agreement and accuracy. One unfortunate historical legacy of Cronbach's (1955) critique was the reframing of questions from assessing judgmental accuracy to examining the process of forming impressions (i.e., from questions of *how well* to questions of *how*—see Gilbert, 1998, p. 92, for a brief review) and almost a complete abandonment of research on interpersonal perceptual accuracy. The relative dearth of research into questions of accuracy and assessing individual differences in accurate interpersonal perception has been attributed to both the opacity of Cronbach's (1955) critique and to the increased data collection effort and analytical sophistication required by his approach. Fortunately, interest and research into basic questions of accuracy in interpersonal perception has revived life into this area (e.g., see Funder, 1987, 1995, 1999; Hall & Bernieri, 2001; Ickes, 1997; Kenny, 1994; McArthur & Baron, 1983; Skowronski & Ambady, 2008; Swann, 1984).

There are two dominant analytical approaches in the post-Cronbach era for assessing accuracy, both of which skirt the original questions of interest. One is to simply bypass the problem that Cronbach identified—trying to assess a single global measure of accuracy and instead rely on correlational methods to estimate the mean level of agreement across individuals and examine potential moderators of the degree of agreement (1955). Such an approach isolates one of Cronbach's components—differential accuracy—to examine individual differences indirectly through the assessment of moderators.[1] For example, levels of self-other agreement are higher for individuals who are more temporally stable (Biesanz & West, 2000; Biesanz, West, & Graziano, 1998).

Another approach has been to reframe the questions of interest. Kenny's social relations model (SRM; Kenny, 1994; Kenny & LaVoie, 1984) provided a framework for documenting levels of consensus in personality and addressing basic interpersonal perceptual questions such as how we see others and are seen by others. Kenny's (1994) SRM has sparked considerable research as it provided a timely and needed methodological framework to examine basic and essential questions of interpersonal perception. This approach allows insight into the general level of agreement (e.g., to what extent is there consensus in perceptions of personality on average across perceivers and targets?) but does not provide a malleable analytical framework for assessing individual differences in accuracy.

Both Cronbach's (1955) components of accuracy model (CCAM) and Kenny's (1994) SRM represent componential approaches to understanding the basic elements of interpersonal perception: Perceivers, Targets, and Measures (see Kenny, West, Malloy, & Albright, 2006, for a review of componential approaches to interpersonal perception). These two models offer different analytical and conceptual lenses for examining interpersonal perception. Cronbach's accuracy components, in conjunction with a set of criteria data, analyze the Target × Measure data separately for each perceiver. In contrast, the SRM analyzes the Perceiver × Target data separately for each Measure. As Kenny, West, et al. (2006, p. 292) note, one cannot conduct both CCAM and SRM modeling simultaneously. Strictly speaking this is absolutely true—each model represents a different analytical path and researchers have faced the dilemma of choosing between these two predominant componential models.

At the same time, the basic questions that drove the original research over half a century ago still cannot be addressed cleanly from these two models. For example, to what extent are there individual differences in perceptive accuracy—the ability to distinguish the characteristics of others? To what extent are there individual differences in expressive accuracy-the level with which an individual is accurately perceived? This article introduces the social accuracy model (SAM) that integrates these two models in a manner designed to address different basic

[1] Ipsatization of measures across a range of different traits is needed to make this fully equivalent to Cronbach's (1955) measure of differential accuracy. See Kenny and Winquist (2001) for detailed examples and more discussion.

questions than those that can be asked from either Cronbach's (1955) accuracy components or Kenny's (1994) SRM. In brief, the SAM is a componential model that estimates perceiver and target effects of Cronbach's components of accuracy. Consequently social accuracy modeling shifts the unit of analysis from the level on a trait, as in a traditional SRM, to the level of accuracy. Instead of modeling how one sees others and is seen by others on a particular trait, SAM examines how accurately one perceives others and is perceived by others *across traits*. SAM provides a third even less traveled path that offers the potential of answering questions of both how accurately we see others and how accurately we are perceived by others as well as offering an analytical framework for elucidating the process of *how* we come to understand others.

This article is organized as follows: First I briefly review Cronbach's (1955) critique and his componential approach. I then provide an overview of Kenny's (1994) SRM. Building on these foundations, I introduce the SAM and provide detailed interpretations of the components of the model. Next I provide several empirical examples interpreting and illustrating the model and its versatility and utility. Discussion focuses on the novel features of the model, the role of componential analyses in understanding interpersonal perception, methodological and analytical issues, and current research using social accuracy modeling that offers new insights into both how and how well we see others.

CRONBACH'S COMPONENTS OF ACCURACY

One historically popular approach for examining individual differences in perceptual accuracy was to take an omnibus measure of the difference between a set of judgments and a corresponding set of criterion measures such as the Euclidean distance—the sum of the squared differences between the two sets of data for each individual. For example, consider a judge, Jordan, who evaluates four targets on three traits (*active*, *reliable*, and *talkative*). These ratings are compared with a validation measure such as self-reports on the same traits (see Figure 1). The sum of the squared differences between the ratings and the validation measures is 34 in this example. This number summarizes the data in Figure 1 into a single global index of the discrepancy between judgments and criteria. Such global summary statistics were often used as a measure of the (*in*)accuracy of a particular perceiver. Cronbach (1955; see also Gage & Cronbach, 1955) noted that multiple different response effects and accuracy components are combined in such a general index and argued instead for partitioning the correspondence between a set of judgments and the validation measures into four components: *elevation accuracy*, *differential elevation accuracy*, *stereotype accuracy*, and *differential accuracy*. These components are derived from the standard 2×2 analysis of variance (ANOVA) decomposition of the target effects, trait effects,

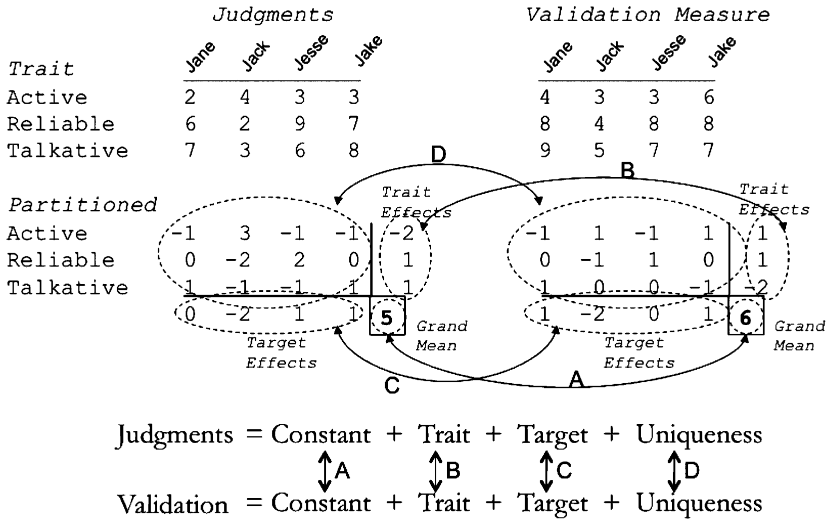

FIGURE 1 Illustration of Cronbach's accuracy components. Judgments represent ratings across four targets on three traits for a single perceiver (Jordan) on a 1- to 9-point scale. Validation measures (e.g., self-reports) are represented for the four targets on the same traits. Partitioned values represent the two-way ANOVA decomposition of each set of data into constant, trait main effect, target main effect, and uniqueness (interaction) values. The relationship between the ANOVA components between judgments and validation measures represents Cronbach's components: (A) elevation accuracy, (B) stereotype accuracy, (C) differential elevation accuracy, and (D) differential accuracy.

and their interaction calculated separately for the set of judgments and the criteria. Thus, as portrayed in the data presented in Figure 1, perceiver Jordan's assessment of how *active* target Jane is (2) in Figure 1 is partitioned into the grand mean of judgments (5), the trait main effect for *active* (−2), the target main effect for Jane (0), and the interaction or uniqueness component (−1): $2 = 5 + (-2) + 0 + (-1)$. Cronbach's (1955) components of accuracy represent the correspondence between these componential effects across judgments and validation measures.

To illustrate using the example presented in Figure 1, *elevation accuracy* is the difference between the grand mean across targets and traits of the judgments ($M = 5$) and the grand mean of the validation measure ($M = 6$; see Path A in Figure 1). *Stereotype accuracy* is the correspondence between the judgments across traits and the "generalized other" or the average person. Stereotype accuracy thus refers to the correspondence between the trait main effects of the judgments (−2, 1, and 1) and the trait main effects of the validation measure (1, 1, and −2) averaged across targets (Path B in Figure 1). *Differential elevation* *accuracy* is the correspondence between the target main effects for judgments and the validation measures. In other words, differential elevation accuracy refers to the correspondence between the four judgment target main effects (0, −2, 1, and 1) and their validation measure target main effects (1, −2, 0, and 1) averaged across traits (Path C in Figure 1). *Differential accuracy* is the relationship between the unique components of judgments and validation measures after removing the trait main effects, the target main effects, and the grand means (Path D in Figure 1). Differential accuracy removes the correspondence attributable to elevation accuracy, differential elevation accuracy, and stereotype accuracy and thus consists of the analysis of the interaction residuals that contain both unique perceiver-target matching effects and measurement error. Following Cronbach (1958), differential accuracy may be calculated separately for each trait that represents the correspondence of simple interaction effects. For example, the differential accuracy in Figure 1 for the trait *active* can be assessed by correlating the unique components of the judgments across targets for that trait (−1, 3, −1, −1) with their corresponding unique components on the validation measure (−1, 1, −1, 1), which results in $r = .58$ for this perceiver (Jordan). For further discussions, illustrations, and examples of Cronbach's (1955) decomposition, see Biesanz and West (2000), Biesanz, West, and Millevoi (2007), Colvin and Bundick (2001), Funder (1999), Kenny (1994), and Kenny and Winquist (2001).

Differential accuracy best captures the ability to discern the unique characteristics of other individuals and is usually assessed by correlating the profile relationship across traits (e.g., Biesanz & West, 2000; Biesanz et al., 2007; Furr, 2008, 2009) or across targets after removing the elevation components and stereotype accuracy. Following Cronbach's (1955) critiques, research into the good judge has generally first calculated differential accuracy correlations or performance on standardized interpersonal perception tests such as the Profile of Nonverbal Sensitivity (e.g., see Hall, 2001) and then in turn assessed the relationship between these correlations and other individual difference measures to identify characteristics of the good judge (for a meta-analytic review see Davis & Kraus, 1997). Such an approach represents a two-stage modeling procedure that neither incorporates measurement error into the analysis nor provides estimates of the extent to which there actually are individual differences in differential accuracy. The inference of individual differences in differential accuracy relies instead on correlations with other individual difference measures. The two-stage modeling approach is inelegant, inefficient, requires a complete and balanced design, and is restrictive in the questions that it allows to be asked. However, through the use of random effects models, the entire analysis can be placed within a single model that will allow a richer set of substantively important questions to be addressed. However, before illustrating such an approach, I review the SRM.

Perceiver	Jane	Jack	Jesse	Jake	Jeb	Perceiver Marginal Means
Jane	-	4	3	3	3	3.63
Jack	6	-	9	7	2	5.50
Jesse	7	3	-	8	3	5.50
Jake	8	2	6	-	4	5.37
Jeb	7	3	6	8	-	5.50
Target Marginal Means	6.63	3.10	6.10	6.57	3.10	5.10

FIGURE 2 Social relations model round-robin design. Data are hypothetical judgments for a single trait (e.g., *active*) on a 1- to 9-point scale. Marginal means for perceivers and targets are adjusted for the round-robin nature of the data (see Kenny, 1994, p. 236) as each perceiver rates a slightly different set of targets.

SOCIAL RELATIONS MODEL

The social relations model (SRM; Kenny, 1994; Kenny & LaVoie, 1984; Warner, Kenny, & Stoto, 1979) is an alternative componential model for examining interpersonal perceptual data that focuses on the perceiver by target data for a single measure. The major components examined within the SRM are perceiver, target, and relationship effects for a measure.[2] To illustrate the SRM, consider the example presented in Figure 2. Figure 2 presents a round-robin design for five persons in which each person rates the other four on a single trait. Each person serves as both a perceiver and a target and the hypothetical data presented are the ratings by each perceiver for each of the other four targets on the trait *active*. Of interest are the marginal means for perceivers and targets—how do perceivers view people on average and how are people seen on average? The SRM allows the partitioning of the variance among ratings into three components—perceiver, target, and, in this example, the combination of specific dyadic variance plus measurement error. With several observations per dyad, measurement error can be separated from unique dyadic variance (relationship variance). In the present example, 18.55% of the variance is attributable to the perceiver effects (termed assimilation in Kenny, 1994) and 35.94% of the variance is attributable to the target effects, reflecting the level of consensus on the trait *active*.[3]

[2]The normal terminology for the SRM refers to perceivers as actors and targets as partners to help emphasize the dyadic nature of the social interaction present in research such as a round-robin design.

[3]Estimation of the variance components was conducted in *R* using the *lmer* package under restricted maximum likelihood.

The SRM provides a nomothetic or generalized approach to assessing questions of interpersonal perception (see Kenny, West, et al., 2006). For instance, the analysis estimates the average level of consensus across the set of perceivers. The analytical framework is malleable, can be adapted to many types of research questions, and has proven extremely useful and productive (for a summary of published research using this model see http://davidakenny.net/srm/srm.htm). Yet, questions of individual differences in levels of consensus or agreement—for either perceivers or targets—cannot be directly addressed within this framework. The classic questions of which perceivers are more accurate (the good judge) and which targets are more accurately perceived (the good or transparent target) require a different analytical framework. The SAM, introduced later, provides an approach to assessing the SRM questions of *average* levels of agreement and consensus while at the same time modeling accuracy and individual differences on these constructs.

THE SOCIAL ACCURACY MODEL

Overview and Definitions

The social accuracy model (SAM; Biesanz, 2007, 2009; for recent empirical examples see Biesanz & Human, 2010; Chan, Rogers, Parisotto, & Biesanz, 2010; Human & Biesanz, 2010; Human, Biesanz, Parisotto, & Dunn, 2010; Lorenzo, Biesanz, & Human, in press) examines the accuracy of a perceiver's impressions of another person (the target). This dyadic unit—the perceptions and impressions that one person has of another—represents the central conceptual level of analysis in the SAM. At the broadest level, *impressionistic accuracy* refers to the accuracy of a perceiver's judgment of the target's personality. Operationally this may be defined as the profile relationship between different assessments of the target and the perceiver's impressions. Impressionistic accuracy is understanding the patterning and ordering of the target's attributes—the holistic impression of that person (e.g., Magnusson & Törestad, 1993). Accuracy in the SAM follows the epistemological framework outlined by Funder (1995, 1999). Specifically, assessing the degree of accuracy is equivalent to determining the validity of a measured construct (Cronbach & Meehl, 1955; Messick, 1989). Multiple different assessments of the target's attributes are needed to help strengthen inferences on the accuracy (validity) of the perceiver impressions. Common validation measures may include target self-reports, reports from close peers and parents, social consensus, and behavioral observations or measurements. Although it is common to examine and report simple correlations to assess the correspondence between perceiver judgments and self-reports (i.e., self-other agreement), in social accuracy modeling the perceiver's judgments are explicitly modeled as the dependent variable with the validation measure(s) serving as independent variable(s). That is, the validity measures—the assessment of the target's personality—is considered fixed and social accuracy modeling assesses the accuracy with which targets are perceived.

Given a high level of impressionistic accuracy for a particular dyadic unit (e.g., Jordan's perception of Jane across the three traits in Figure 1), how can one interpret this high level? Is Jordan generally very accurate in her perceptions of others and thus high in *perceptive accuracy*—the extent to which a particular perceiver's impressions are more or less accurate than other perceivers on average across different targets? Perceptive accuracy is an assessment of the extent to which someone is a good judge of others. A separate question is the degree to which Jane is accurately perceived by others and thus high in *expressive accuracy*—the extent to which a particular target is accurately perceived on average across different perceivers.[4] Expressive accuracy is an assessment of the extent to which someone is a good target and has been called readable, legible, judgable, and transparent.

Understanding and interpreting levels of impressionistic accuracy requires more than just ascertaining the perceiver and target accuracy effects of perceptive and expressive accuracy. In general, overall levels of impressionistic accuracy are expected to be quite substantial. Jordan's impression of Jane across a large number of attributes is expected to be associated with the impression that any other perceiver forms of another randomly chosen person. Blackman and Funder (1998) provided an empirical example of this phenomenon and demonstrated that the impression of different perceivers who formed impressions of *different* targets correlated on average at $r = .18$ to .26 across experimental conditions. This effect arises when there are mean-level differences—averaged across targets—on the validation measures. Ratings and self-reports on an item such as "*Is outgoing, sociable*" are generally substantially higher than on "*Starts quarrels with others*." Thus perceiver judgments for *different* targets are expected to be moderately and positively related across personality traits because different targets are indeed similar on the ordering and patterning of personality attributes on average. Thus it would not be surprising to observe the impressions of different perceivers evaluating two different targets to be associated across measures. Levels of impressionistic accuracy could be high simply from knowledge of what the average (normative) person is like. Measures of impressionistic accuracy—particularly for single perceiver-target dyads—suffer from many of the same criticisms raised by Cronbach (1955) and consequently impressionistic accuracy will typically *not* be of primary interest.

[4]See Human (2009) for the development and more extensive discussion of the terms *perceptive* and *expressive accuracy*.

Instead, the focus should not only be global accuracy—the levels of accuracy for perceivers across different targets and for targets across different perceivers—but also on specific components of accuracy. The SAM decomposes impressionistic accuracy into assessments of individual differences in stereotype accuracy and differential accuracy. Following Furr (2008) I also refer to these assessments as normative and distinctive accuracy, respectively. The normative and distinctive accuracy components are modeled for both perceivers and targets (see Figure 4). Before introducing the general model, I first examine in detail the basic elements that comprise the SAM.

Estimating the Social Accuracy Model

The basic unit in the SAM is a perceiver's judgment of a target across a series of attributes or traits. This unit is predicted by a measure used to validate the perceiver's impressions. This profile relationship provides an index of how well the perceiver's judgments correspond with the validation measure. The validation measure may be a target's self-report on those same measures, a knowledgeable informant report, behavioral observations, or a composite of these different validation measures. This simple profile relationship is expressed in the following linear regression equation:

$$Y_k = \beta_0 + \beta_1 V_k + \varepsilon_k. \tag{1}$$

The relationship here is expressed as an *unstandardized* regression equation with impressions as the dependent variable across k different personality attributes. This contrasts to the more common correlational metric to highlight the importance of modeling the perceiver's impressions—the importance of which will become more apparent as the components of the perceiver's impressions are elucidated and developed shortly. Thus β_1 refers to the population unstandardized regression coefficient relating the target validity measures (V_k; e.g., self-reports or informant reports) to the perceiver's impressions (Y_k) that is estimated by unstandardized regression coefficient b_1. This represents the average relationship between the target validity measures and perceiver ratings across traits. This relationship may vary across traits, which is discussed later; for now the focus is on the main effect of the average relationship across traits as this will allow the estimation of individual differences of target accuracy effects, which is not feasible for single trait. The SAM assumes that the validation measure's scale is constant across different personality attributes. Combining different measurement instruments in the same analysis would present substantial conceptual and interpretational difficulties. Subscripts for the perceiver and target are omitted in the interest of simplicity as the present focus is on a single perceiver-target

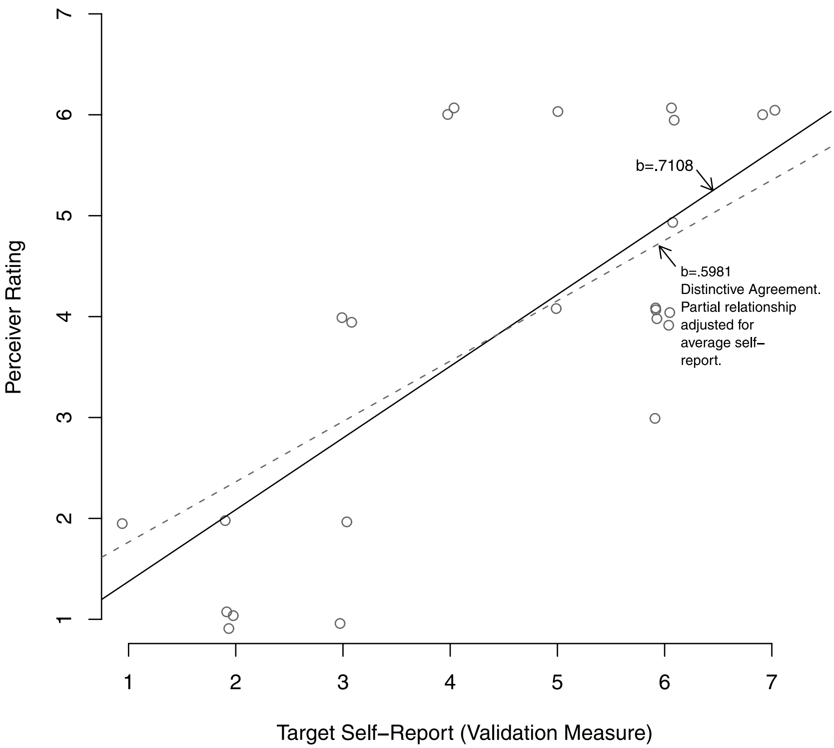

FIGURE 3a Illustration of raw profile agreement (solid line; see Equation (1)) and distinctive accuracy after partialling the normative self-report response profile (dashed line; see Equation (2)).

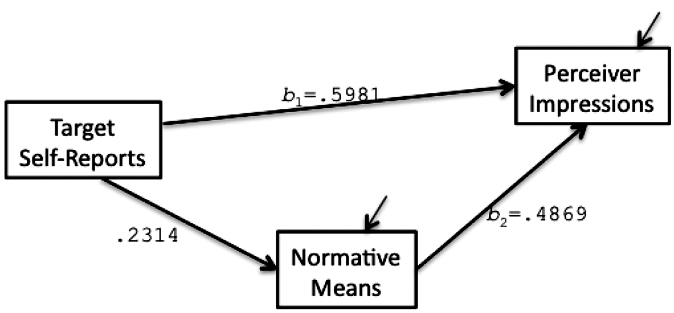

FIGURE 3b Illustration of the mathematical mediational decomposition for a single perceiver-target dyad.

	Perceiver *Perceptive accuracy*	Target *Expressive accuracy*
Distinctive Accuracy	The extent to which one perceives the distinct, unique characteristics of others.	The extent to which one's unique and distinct characteristics are perceived by others.
Normative Accuracy	How much one's impressions of others corresponds to that of the average person.	How similar to the average person one is generally perceived.

FIGURE 4 Definition of the four main social accuracy model random effects. Intercept and dyadic random effects may also be estimated within the model (see Table 1).

pair. Positive values of b_1 indicate that the perceiver's impressions are related to the validation measure.

Consider the following set of data illustrated in Figure 3a for a single perceiver-target dyad where $b_1 = .7108$ across 24 personality trait measures with the target self-reports serving as the validation measure.[5] This relationship indicates a strong correspondence between the perceiver's impressions and the target's self-reports. How are we to understand this level of self-other profile agreement?

In Equation (1) β_1 represents the population level of impressionistic accuracy, which is a blend of distinctive (differential) agreement and normative (stereotype) agreement. Cronbach's (1955) terminology is noted in parentheses. Consequently, to understand, interpret, and more appropriately model impressions entails expanding Equation (1) to include the relationship with the generalized average person. This is accomplished by including the additional predictor μ_k, the average target standing on each measure or trait for that validation measure, as well as additional subscripts i and j to denote perceiver and target, respectively, as follows:

$$Y_{ijk} = \beta_{0ij} + \beta_{1ij} V_{jk} + \beta_{2ij} \mu_k + \varepsilon_{ijk}. \tag{2}$$

The partial regression coefficients in Equation (2) provide components analogous to Cronbach's (1955) accuracy components. The coefficient β_{1ij} now represents a measure of differential or distinctive accuracy—to what extent do the perceiver's impressions relate to the validation measure after partialling out the average per-

[5]Estimates in this example are carried to four decimal places in order to illustrate several mathematical relationships.

son's standing on the validation measure? The partial regression coefficient β_{2ij} represents a measure of stereotype accuracy—to what extent are the perceiver's impressions related to the generalized other based on the validation measure?

The mean value of the validity measure, μ_k, is a parameter and cannot be determined exactly. In practice, an estimate of this expected value, $\hat{\mu}_k = \overline{V}_k$, such as the sample mean based on the validation measures (e.g., the mean self-report here for each personality measure), is used in Equation (2). If the study in question has only a small or moderate sample size, but if there is additional data from this population that provides more precise estimates of $\hat{\mu}_k$, the use of these more precise estimates is warranted as long as the participants in the current study can be reasonably considered exchangeable with those from the larger set of data. In the present example, $b_1 = 0.5981$ and $b_2 = 0.4869$ indicating both substantial distinctive and normative agreement for this particular perceiver-target pair.

Cronbach's (1955) original formulation defined stereotype accuracy as the relationship between perceivers' generalized impressions (averaged across targets) with the generalized validation measures (again averaged across targets). In the SAM this association is defined after partialling out the target's actual standing on the validation measure and assessed for each perceiver-target dyad. Thus, this measure, although conceptually analogous to Cronbach's formulation, is defined slightly differently. Defining distinctive and normative accuracy as in Equation (2) has benefits: (a) it allows a full conceptual decomposition of the raw profile relationship (impressionistic accuracy) and (b) these components can also be modeled for both perceivers and targets.

Examining Components of Accuracy: Estimating Distinctive and Normative Accuracy

Traditional Cronbachian accuracy components are not additive in the sense of allowing one to recreate the raw profile relationship (see also Furr, 2008). However, the basic expression of the SAM in Equation (2) allows the separation of overall profile agreement into the distinctive component and the normative mediated component (see Figure 3b). In the present example, the relationship between the target's self-report and the perceiver's impressions is exactly the sum of the distinctive component ($b_1 = 0.5981$) and the relationship between the perceiver's impressions and the self-report that is mediated through the generalized other—the normative self-report profile. This mediated component is the product of the relationship of the target's self-report profile with the normative self-report profile ($b = 0.2314$) with the normative coefficient from Equation (2), $b_2 = 0.4869$. The mediated normative component is therefore $0.2314 \times 0.4869 = 0.1127$. The total profile relationship (self-other agreement across the 24 personality ratings) for this perceiver-target pair of .7108 is the sum of the distinctive component (0.5981) and the normative mediated component (0.1127). This relationship holds exactly for a single perceiver-target dyad.

The use of the mediational framework provides a useful mathematical tool for the profile decomposition—this relationship holds exactly for linear regression models with no missing data. Strictly speaking, however, this is not a true theoretical mediational model in the sense that the target's self-report "causes" the normative response profile, μ_k (see MacKinnon, Krull, & Lockwood, 2000, for discussion of this decomposition when one is not positing causal mediation). This mathematical decomposition is better viewed instead as a useful theoretical vehicle for understanding the different components of profile agreement and separating perceiver's impressions into distinctive and normative components. Impressionistic accuracy is the sum of distinctive accuracy and the normative mediated component.[6]

In general, accuracy estimates for specific perceiver-target dyads will be of limited interest. Interpreting the distinctive component ($b_1 = 0.5981$) is difficult—is this value high because this perceiver is high in perceptive accuracy or because this target is generally very accurately perceived and thus high in expressive accuracy? Estimates of perceptive accuracy across different targets and estimates of expressive accuracy across different perceivers are needed instead. Only research designs where there are multiple perceivers and multiple targets can provide estimates of individual differences attributable to perceivers and targets. To illustrate the model, how to estimate SAM, and interpret the resulting model, Study 1 presents and interprets a half-block design (classic design; see Kenny & Albright, 1987) and Study 2 a round-robin design.

STUDY 1

Method

A total of 202 undergraduates (145 females; mean age = 18.80 years, SD = 2.11) participated in exchange for course credit. All participants were individually provided with a description of the mechanics of the experiment followed by the same verbal instructions: "*You will first fill out a short questionnaire about yourself and then watch a series of seven short interviews on the computer listening to them on headphones. For each interview, we are interested in your impressions of that person's personality. So after each interview there will be*

[6]Although distinctive accuracy can be estimated for a single perceiver-target dyad, normative accuracy is confounded with the target's actual normativeness. Estimating a single perceiver's normative accuracy requires determining the normativeness of that perceiver's impressions on average across a large number of targets. In other words, how does the perceiver's impressions of others, on average, correspond to the average validity measure?

a short questionnaire about your impressions of that person's personality." Participants then watched videotapes of a common set of seven undergraduate women each answering basic "getting-to-know-you" questions for approximately 5 min posed by a common undergraduate female interviewer. After each video was presented, personality impressions of each target were assessed using the 44-item Big Five Inventory, which covers a diverse range of core personality characteristics (BFI; see John & Srivastava, 1999) on a scale from 1 (*disagree strongly*) to 9 (*agree strongly*). All videos were presented in the same order. Target self-reports on the BFI were used to validate perceivers' impressions.

Analytical Methods, Results, and Interpretation

This experiment represents a classic design in which the same set of seven targets are rated by a large ($n = 202$) number of perceivers on the same 44 personality items. Modeling self-other agreement across the 44 items in the BFI, after accounting for perceiver and target random effects, entails a crossed-random effects analysis that requires multilevel modeling. The specific equations for this model are as follows:

$$Y_{ijk} = \beta_{0ij} + \beta_{1ij} TSelf_{jk} + \beta_{2ij} Mean_k + \varepsilon_{ijk} \tag{2.1}$$

$$\begin{aligned} \beta_{0ij} &= \beta_{00} + u_{0i} + u_{0j} + u_{0(ij)} \\ \beta_{1ij} &= \beta_{10} + u_{1i} + u_{1j} + u_{1(ij)} \\ \beta_{2ij} &= \beta_{20} + u_{2i} + u_{2j} + u_{2(ij)}. \end{aligned} \tag{2.2}$$

Here Y_{ijk} is perceiver i's rating of target j on item k; $TSelf_{jk}$ is target j's self-report on item k (i.e., the validation measure V_{jk}). The predictor $Mean_k$ is the mean self-report on item k based on a larger set of self-reports ($n = 1{,}157$) from similar participants in other studies (i.e., the estimated average validation measure response for each item $\hat{\mu}_k$) providing a more stable estimate of this parameter. For the perceiver i–target j dyad, the estimated regression coefficient $\hat{\beta}_{0ij}$ is the intercept (predicted rating when $TSelf_{jk} = 0$ and $Mean_k = 0$). The estimated coefficient $\hat{\beta}_{1ij}$ represents an estimate of distinctive self-other agreement for perceiver i with target j–the level of self-other agreement after holding constant $Mean_k$. In other words, $\hat{\beta}_{1ij}$ is the estimated level of self-other agreement holding constant and controlling for the average person's self-reported personality profile. Finally, $\hat{\beta}_{2ij}$ is the estimated level of normative agreement for perceiver i with target j–the correspondence between the perceiver's ratings and the average person's self-reported profile after partialling target j's self-report.

In the present data set, Equation (2.1), the Level 1 equation, is estimated for each perceiver-target dyad resulting in $202 \times 7 = 1{,}414$ regression equations across 44 personality items. The observed variability across these 1,414 sets of regression coefficients is the sum of the true latent variability (individual differences in accuracy components) and random sampling fluctuations. Conceptually, Equation (2.2), the upper level equations, performs a standard two-way ANOVA on the true latent variability separating these random effects into the perceiver main effects, the target main effects, and the residual dyadic components that represent unique effects plus measurement error.

The fixed effects $\{\beta_{00}, \beta_{10}, \beta_{20}\}$ represent the average intercept, distinctive accuracy, and normative slope, respectively, across both perceivers and targets. The u's represent the random (latent) effects in terms of deviations from the grand mean (fixed effects). For instance, u_{1i} is perceiver i's unique distinctive accuracy slope averaged across the seven targets, u_{1j} is target j's unique distinctive accuracy slope averaged across the 202 perceivers, and $u_{1(ij)}$ is the dyadic plus residual component—the accuracy of perceiver i evaluating target j after removing the grand mean (β_{10}) and the perceiver and target main effects (u_{1i} and u_{1j}). Thus $\beta_{10} + u_{1i}$ represents perceiver i's level of distinctive self-other agreement averaged across the seven targets; $\beta_{20} + u_{2j}$ represents target j's normative self-other agreement averaged across the 202 perceivers. Specific latent values are not uniquely estimated—instead the variance of u_{1i} across perceivers and u_{1j} across targets is estimated. This model can be estimated using any standard multilevel modeling software program with three grouping variables: perceiver, target, and the specific perceiver-target dyad. All results presented were estimated using *R*'s *lme4* package (Bates & Sarkar, 2007; *R* Development Core Team, 2009). I discuss the results and interpretation of the base model illustrated in Equations (2.1) and (2.2) before examining how the perceiver's personality moderates distinctive and normative accuracy.

Distinctive accuracy. On average across perceivers and targets, there was significant self-other agreement across the 44 items capturing the Big Five personality factors after controlling for the normative response profile, $b = 0.26$, $z = 3.91$, $p = .0001$ (see Table 1). Perceiver ratings evidenced agreement with target self-reports on their unique and individuating attributes. Examination of the random effects revealed relatively few individual differences among perceivers, estimated *SD* for $u_{1i} = .07$. This finding is consistent with Kenny's (1994) conclusion based on a review of the research on the good judge that assessments of the good judge are generally not made reliably. Reliability may be assessed as true score variance (individual differences in the good judge) divided by true score variance plus error variance. This calculation results in a fairly low value in the present study due to the small variance among the true scores. In contrast, there were more individual differences among the seven

TABLE 1
Fixed Effect Estimates From the Social Accuracy Model

Fixed Effect	*Study 1* (n = *202)* *Estimate* (SE)	*Study 2* (n = *273)* *Estimate* (SE)
Fixed Effects (Base Model)		
Intercept (b_0)	1.86 (0.67)**	−0.03 (0.12)
Target self-report (b_1)	0.26 (.07)**	0.08 (0.01)**
Normative profile (b_2)	0.35 (.17)*	0.87 (0.03)**
Perceiver Personality Moderation Effects		
Target self-report (Distinctive accuracy, b_{11})		
Agreeableness	0.00 (0.01)	0.00 (0.01)
Conscientiousness	0.01 (0.01)	0.00 (0.01)
Extraversion	0.00 (0.01)	−0.00 (0.01)
Neuroticism	0.01 (0.00)	0.00 (0.00)
Openness	0.02 (0.01)**	0.02 (0.01)**
Normative Profile (Normative Accuracy, b_{21})		
Agreeableness	0.08 (0.02)**	0.16 (0.03)**
Conscientiousness	0.05 (0.02)**	0.08 (0.03)**
Extraversion	0.06 (0.02)**	0.05 (0.02)*
Neuroticism	−0.03 (0.02)*	−0.09 (0.02)**
Openness	0.02 (0.02)	0.05 (0.03)

Note. Base model is without any interactions and corresponds with Equation (2.2) and Table 1. Interactions were estimated separately for each personality trait for both distinctive and normative accuracy simultaneously (e.g., see Equation (2.3)).

$*p < .05$. $**p < .01$.

targets, estimated *SD* for $u_j = .18$ (see Table 2). Making strong comparisons between these two estimates is difficult in the present design as estimates of variability based on seven target observations are not precise relative to those based on 202 perceiver observations.

Distinctive accuracy has two equivalent interpretations. First, $\beta_{1i} = \beta_{10} + u_{1i}$ is the average self-other profile agreement for perceiver i after partialling out normative self-reported levels on that item. In other words, in this experiment each perceiver has seven profile relationships across 44 items (one for each target). β_{1i} represents the average of these seven profile relationships after controlling for the mean level. Second, it is *also* equivalent to the average level of self-other agreement across the seven targets computed separately for each item. Specifically, if target self-reports are centered within target and self-other agreement across the seven targets is computed for perceiver i separately for each of the 44 items, then β_{1i} represents the average of these levels of self-other agreement across the 44 relationships. For a more thorough discussion of these nonintuitive equivalent interpretations, see Kenny and Winquist (2001, pp. 275–278) and Biesanz and Human (2010).

TABLE 2
Random Effect Estimates From the Social Accuracy Model

Random Effects	*Study 1* (n = *202)* SD *Estimate*	*Study 2* (n = *273)* SD *Estimate*
Perceiver (Perceptive Accuracy)		
Intercept (u_{0_i})	1.66	1.60
Target self-report (u_{1_i})	0.07	0.01
Normative profile (u_{2_i})	0.28	0.36
Target (Expressive Accuracy)		
Intercept (u_{0_j})	1.74	0.98
Target self-report (u_{1_j})	0.18	0.20
Normative profile (u_{2_j})	0.45	0.32
Dyadic Variability		
Intercept ($u_{0_{(ij)}}$)	1.69	0.00
Target self-report ($u_{1_{(ij)}}$)	0.01	0.11
Normative profile ($u_{2_{(ij)}}$)	0.30	0.12
Residual *SD*	1.65	1.17

Note. Random effects estimates are from the base model that corresponds to Equation (2.2) with no interactions.

To understand how these two different analyses are equivalent in the present case, consider a hypothetical perceiver, say Jordan, who rates several targets with distinctive self-other profile agreement (e.g., $b_{1i} = .40$). Jordan has a moderate ability to discern whether Jack self-reports higher or lower on "*Is outgoing, sociable*" than the average person does on this trait as well as whether he self-reports higher or lower on "*Starts quarrels with others*" than the average person. For each item Jordan can predict the extent to which someone will self-report higher or lower than the average person will on that specific item. This ability to determine if a specific target will self-report higher or lower than the average person is exactly what is measured examining self-other agreement on a single trait across targets. If Jordan can predict whether Jack is higher or lower on "*Is outgoing, sociable*" than the average person, she has that level of ability to determine whether Jane is higher or lower than the average person on that same trait and thus if Jack is higher or lower than Jane on that trait.

In more abstract terms, each analysis is based on exactly the same set of validation data (44 items by 7 targets = 308 self-reports). This predictor data is a matrix (44 rows by 7 columns). The profile analysis consists of analyzing each column separately and then averaging the seven profile relationships. In contrast, the trait-level analysis analyzes each row separately and then averages the 44 item-level relationships. Each method provides a summary of the total level of self-other (distinctive) agreement across all 308 validation data points, which is what is accomplished in Model 2. More formally, if the variances of the items are homogeneous, as the number of targets and items increase, both analyses will converge to the same value. In finite and small samples there may be slight numerical differences in these two approaches due to sampling fluctuations when estimating and combining different regression equations.

Normative accuracy. Across perceivers and targets the normative self-reported response profile was related to perceiver ratings after adjusting for target self-reports, $b = .35$, $z = 2.02$, $p = .04$. As well, there were substantial individual differences among perceivers in their normative accuracy coefficient, estimated *SD* for $u_{2i} = .28$, as well as among targets, estimated *SD* for $u_{2j} = .45$.

The interpretation of the normative accuracy relationship has multiple features. First, responses to the 44 items analyzed in Equation (2) are *not* reverse coded. As a consequence there is a substantial relationship between $Mean_k$ and social desirability (see Borkenau & Zaltauskas, 2009; Edwards, 1957). Indeed, across the 44 items on the BFI the mean self-report on an item ($Mean_k$) correlates extremely strongly with the average social desirability of that item based on a separate sample ($n = 486$; Paulhus, 2009), $r(42) = .86$, $p < .00001$. Higher levels of normative accuracy are thus associated with social desirability and more positive impressions (Wood, Gosling, & Potter, 2007). Indeed, replacing $Mean_k$ with the average social desirability rating for that item in Equation (2.1) yields results similar to those presented in Tables 1 and 2; however, such an analysis no longer provides measures that can be clearly interpreted as distinctive and normative self-other agreement. These results suggest that normative accuracy may be largely based on a strong evaluative component in many data sets.

Second, perceptive normative accuracy (e.g., $\beta_{2i} = \beta_{20} + u_{2i}$) represents how a given perceiver evaluates targets *on average*. Critical here is the expected relationship between the validation measure and the normative validation measure. Under some minimal assumptions this expected unstandardized relationship is actually the intraclass correlation.[7] For every perceiver the expected average relationship between target self-reports and the normative self-report response

[7]Define $\beta_{VN} = E(b_{VN_j})$ as the expected unstandardized regression equation predicting the target's validity measures from the normative validity measures (i.e., $\hat{V}_{jk} = b_{0j} + b_{VN_j}\mu_k$) and $\beta_{NV} = E(b_{NV_j})$ as the reciprocal relationship (i.e., $\hat{\mu}_k = b_{0j} + b_{NV_j}V_{jk}$). All expectations are taken across targets. Because $E(\hat{V}_{jk}) = \mu_k$, $\beta_{VN} = E(b_{VN_j}) = E(\frac{s_{V_{kj}}}{s_{\mu_k}} r_{V_{kj}\mu_k}) = 1$. In other words, the expected unstandardized regression coefficient predicting the validity measure from the average response on that validity measure across k measures is exactly 1.00. Consequently the expected unstandardized relationship between the validity measure predicting the average validity measure will be $\beta_{NV} = E(b_{NV_j}) = E(\frac{s_{\mu_k}}{s_{V_{kj}}} r_{V_{kj}\mu_k}) = E(r_{V_{kj}\mu_k})^2$. The expected unstandardized relationship between the target's validity measure and the normative profile is, asymptotically, the intraclass correlation coefficient.

profile ($Mean_k$) across targets is this same constant value. Consequently higher values of β_{2i} represent (a) a stronger relationship between normative profile and the perceiver's impressions relative to a perceiver with a lower value of β_{2i} and (b) greater generalizability of the perceiver's impressions across targets. Individual differences among perceivers in normative accuracy may reflect a combination of evaluative tendencies and actual generalized knowledge of personality—what individuals are actually like on average.

Similar logic is used for interpreting expressive normative accuracy. Here $\beta_{2j} = \beta_{20} + u_{2j}$ represents the average normative coefficient for a given target j averaged across different perceivers. Because targets will vary in the normativeness of their personalities, interpretation of β_{2j} requires consideration of the actual relationship between the target's self-report and the average self-reported personality profile. This relationship may vary substantially across targets. Consequently if target characteristics are examined as potential moderators, it may be necessary to compute the mediated relationship of the validation measure through the normative validation measure. This represents a multilevel mediational model, which, in this study, requires estimation of the covariance across targets between (a) the paths between the target validity measure and the normative profile and (b) the normative coefficient u_{2j} from Equation (2.3). For more discussion of examining moderated mediation within multilevel models see Kenny, Korchmaros, and Bolger (2003) and Bauer, Preacher, and Gil (2006).

Examining potential moderators. Equation (2) can be expanded to include potential moderators of components of accuracy. This study examines the relationship between the perceiver's personality as assessed by Big Five trait levels on the BFI and normative and distinctive accuracy. To illustrate with Agreeableness, denoted as (AG_i), Equation (2.2) is expanded as follows:

$$\begin{aligned} \beta_{0ij} &= \beta_{00} + \beta_{01}AG_i + u_{0i} + u_{0j} + u_{0(ij)} \\ \beta_{1ij} &= \beta_{10} + \beta_{11}AG_i + u_{1i} + u_{1j} + u_{1(ij)} \\ \beta_{2ij} &= \beta_{20} + \beta_{21}AG_i + u_{2i} + u_{2j} + u_{2(ij)}. \end{aligned} \tag{2.3}$$

Here AG_i is examined as a potential moderator for distinctive accuracy (β_{11}) as well as normative accuracy (β_{21}). Perceiver Agreeableness was significantly associated with higher levels of perceptive normative accuracy, $b = .08$, $z = 4.07$, $p = .00005$, but not with distinctive accuracy, $b = .00$, $z = 0.72$, $p = .47$. Parallel results to Study 1 were obtained with perceiver Neuroticism, Conscientiousness, and Extraversion (see Table 2). Interestingly, the opposite pattern was observed for Openness, which was positively associated with higher levels of distinctive accuracy, $b = .02$, $z = 2.86$, $p = .004$, but not normative accuracy, $b = .02$, $z = 0.99$, $p = .32$.

The primary limitation with a classic Perceivers × Targets design such as Study 1 is that estimates related to individual differences with respect to targets are limited by the constraints on the design—it is difficult to have perceivers assess a large number of different targets. Alternative designs such as round-robin design where perceivers are also targets overcomes this limitation when there are a large number of groups. Study 2 examines the SAM within a round-robin design and examines the same analyses while deferring discussion of the technical modeling differences.

STUDY 2

Method

A total of 273 undergraduates (199 females, mean age = 20.90, *SD* = 4.16) at the University of British Columbia participated in 45 groups, ranging in size from 3 to 12 (Median = 7), in exchange for $20 or two extra course credits. Participants engaged in a round-robin "getting-acquainted" design within each group. After self-assessments of their own personalities, participants paired up and met with another group member for 3 min in an unstructured interaction before separating to provide their impressions of the other participant's personality. This process was repeated until all participants had met and provided impressions of every other participant in their group. Self-reports and ratings were assessed using a 21-item abbreviated version of the BFI with the inclusion of 3 additional items to assess intelligence: "*Is intelligent,*" "*Is bright,*" and "*Receives good grades.*" All items were assessed on a 1 (*disagree strongly*) to 7 (*agree strongly*) rating scale. The specific BFI items assessed are detailed in Human (2009), Human and Biesanz (2010), and available from Jeremy C. Biesanz. Perceivers provided self-reports on the full 44-item BFI.

Analytical Methods and Results

The analytical model in the context of a round-robin analysis is the same as in the classic half-block design and expressed in Equations (2.1–2.3). However, notable technical differences result from the round-robin design. Participants are both perceivers and targets. As a consequence of this cross-classification, there are as many targets as perceivers allowing equally precise estimates of the heterogeneity in the accuracy component random effects for both perceivers and targets.

Distinctive accuracy. On average across perceivers and targets, there was substantial self-other agreement across the 24 items capturing the Big Five personality factors after controlling for the normative response profile, $b = 0.08$, $z = 6.21$, $p < .0001$ (see Table 2). Again, perceiver ratings evidenced substantial agreement with target self-reports on their unique and individuating attributes. Similar to Study 1, estimation of the random effects reveals relatively few individual differences among perceivers, estimated *SD* for $u_{1i} = .01$. In contrast, there was substantially more variability across targets in their levels of expressive distinctive accuracy, estimated *SD* for $u_{1j} = .20$.

Normative accuracy. Across perceivers and targets the normative self-reported response profile was strongly related to perceiver ratings after adjusting for target self-reports, $b = .87$, $z = 27.69$, $p < .0001$. As well, there were substantial individual differences among perceivers in their normative accuracy coefficient, estimated *SD* for $u_{2i} = .36$, as well as among targets, estimated *SD* for $u_{2j} = .32$. For perceivers, individual differences in the normative accuracy component may reflect a blend of individual differences in generalized knowledge (i.e., knowledge of what the average person is actually like) coupled with a generalized evaluative tendency. In contrast, how normatively targets are perceived reflects a generalized evaluation across perceivers coupled with the target's actual normativeness.

Perceiver personality as a moderator of accuracy components. Equation (2.3) was examined for each of the perceiver's Big Five trait levels as a potential moderator of the components of accuracy. Perceiver Agreeableness was again significantly associated with higher levels of perceptive normative accuracy, $b = .16$, $z = 5.35$, $p < .0001$, but not with distinctive accuracy, $b = .00$, $z = 0.09$, $p = .93$. Parallel results were obtained with perceiver Neuroticism, Conscientiousness, and Extraversion. The results observed in Study 1 for Openness replicated in Study 2 as higher levels of Openness were again associated with greater distinctive accuracy, $b = .02$, $z = 2.74$, $p = .006$, but not significantly with normative accuracy, $b = .05$, $z = 1.79$, $p = .07$. Overall, the exact pattern of results observed in Study 1 was replicated in Study 2.

DISCUSSION

Across both studies, target self-reports were significantly related to perceiver impressions both distinctively and normatively. This replicates substantial previous research examining the accuracy of initial impressions of personality. At the same time, the SAM provides a wealth of additional information that can be illuminating. One of the clear conclusions from the SAM in Studies 1 and 2 is that the classic good judge of personality—perceiver distinctive accuracy—is precisely the component of accuracy where there does not appear to be very reliable individual differences, at least in the contexts examined in these studies. For instance, Figure 5 clearly shows that individual differences in the classic good judge are relatively scant—there is relatively little heterogeneity across perceivers in their ability to detect the unique and individuating personality characteristics of others. In contrast, the three other main effects estimated by the SAM—target distinctive accuracy and perceiver and target normative accuracy for both perceivers and targets all evidenced substantial reliable variability. The four random effects estimated in Study 2 are defined in Figure 4 and their distributions are illustrated in Figure 5. This suggests that systematic research

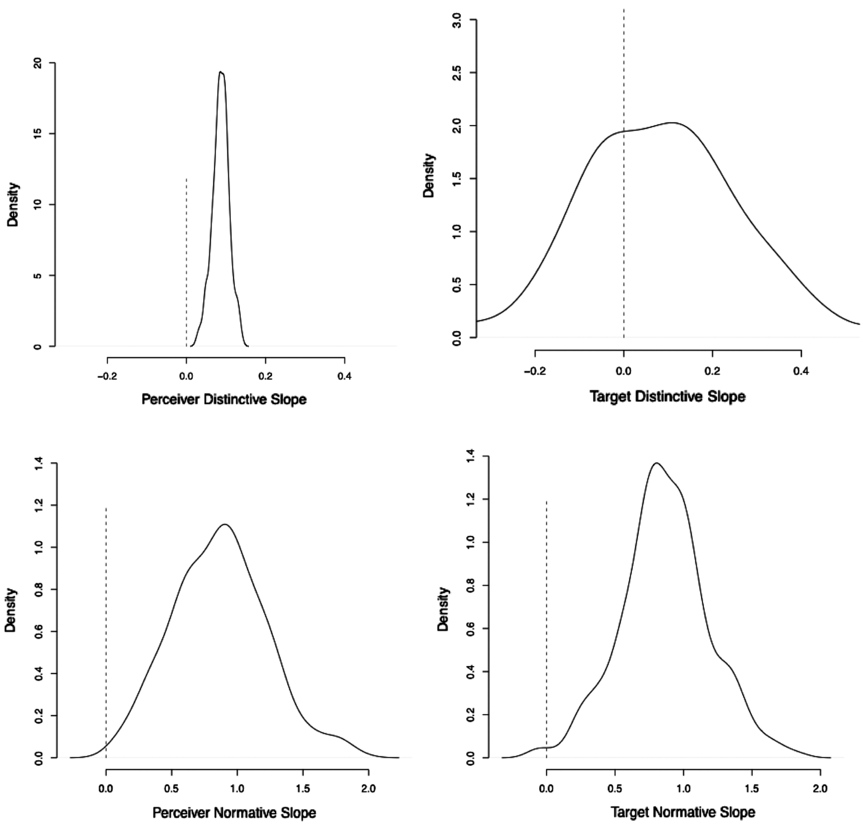

FIGURE 5 Kernel density plots portraying the distribution of the random effects corresponding to the four main social accuracy model effects in Study 2 ($n = 273$). See Figure 4 for definitions. Note that the graphed distribution of the random effects corresponds to the estimated mean and variance from the random effects model in Tables 1 and 2 (e.g., see McDonald, 1981; Shen & Louis, 1999).

examining the good target—what are the characteristics of individuals who are more accurately perceived (e.g., Biesanz & West, 2000; Biesanz, West, & Graziano, 1998; Colvin, 1993a, 1993b)—is worth systematically exploring. Indeed, preliminary research indicates that target adjustment is broadly related to being accurately perceived (Human, 2009).

Modeling Issues in the Social Accuracy Model

The analyses presented for Studies 1 and 2 can be recreated using any standard multilevel modeling software (e.g., *R*, PASW/SPSS Mixed, STATA, MLwiN, SAS Proc Mixed, HLM, etc.). However, there are a number of modeling issues that may arise. For instance, it may be useful to relax the homogeneity of variance constraint and allow heterogeneity in the variance of the errors in Equation (2.1) across items (e.g., different items may have different error variances). This is not presently possible in all software packages. In the round-robin design there may be heterogeneity associated with the g different groups, which can be estimated through the inclusion of another set of random effects in Equation (2.2)—that is, estimating the variance associated with $\{u_{0g}, u_{1g}, u_{2g}\}$. In addition there may be associations within the dyads such that the distinctive accuracy with which a perceiver i sees target j is correlated with how that perceiver j sees target i. Positive correlations within dyad can be modeled within standard multilevel software through the inclusion of the (i, j) pairing as an additional random effect.

The round-robin design in particular presents an additional unique set of challenges in estimation. For precisely this reason Kenny (1998) developed the SOREMO package for the SRM. Standard multilevel software estimating cross-classified random effects assumes independence of observations across different perceiver-target cells. This assumption is met for the classic Judge $\times$ Target half-block design when the set of perceivers are different participants than the targets. Consequently the random effects for perceivers are independent of the random effects for targets. In the round-robin design participants serve as *both* perceivers and targets. It is quite possible that there is an association between how accurately one perceives others and how accurately one is perceived by others. The covariances between perceiver effects and target effects are not estimated within standard software. What then are the options for estimating such effects that may arise in round-robin data?

One approach would be to use Bayesian Markov chain Monte Carlo (MCMC) modeling. Gill and Swartz (2001, 2007) present an approach for estimating the SRM using MCMC modeling (see also Wong, 1982, for expectation-maximization estimation routines). Browne, Goldstein, and Rasbash (2001) present MCMC modeling for cross-classified units. These basic approaches can be extended to multivariate data with predictors such as Equation (2). However, regardless of its theoretical elegance, MCMC modeling in the present context currently presents serious analytical and computational barriers as it would require substantial expertise to appropriately program, implement, and interpret.

More important, it is not clear that explicitly estimating the perceiver-target covariances is necessary. To elaborate, although standard multilevel software assumes that cross-classified effects are independent, this is not an actual constraint placed on the estimation. These covariances are not estimated as opposed to being constrained to zero. Thus if a perceiver's distinctive accuracy is correlated with his or her target distinctive accuracy effect, this association will be present when examining the random effects for perceiver and target distinctive accuracy. Standard multilevel modeling software currently simply treats perceivers and targets as being different sets of participants. At a practical level, the question is what impact does using standard software to estimate the SAM based on round-robin data have on parameter estimates and inferences? Following Steiger's (2006a, 2006b) exhortation to simulate the impact of modeling assumptions within the context of a specific data set, I conducted a brief Monte Carlo simulation study based on the exact design characteristics in Study 2. Specifically, the target self-report predictor matrix in Study 2 was used for the simulation study and parameters examined were consistent with those estimated. Three levels of perceiver-target random effect correlations were examined ($\rho_{pt} = 0$, .30, and .50) and 3,000 simulations were conducted for each level. Table 3 presents the parameter values and the results of the simulation. When $\rho_{pt} = 0$, the assumption of independence between the perceiver-target factors is met. Interestingly, moderate ($\rho_{pt} = 0.30$) and large ($\rho_{pt} = 0.50$) correlations between perceiver-target random effects had no discernible impact on estimation of random effects, fixed effects, or standard errors of the fixed effects. This suggests that, at least for the present design and in the parameter space corresponding to the observed estimates, not modeling the perceiver-target covariance has no impact on the resulting analysis. Thus in Study 2, the use of standard multilevel software appears to be appropriate and justified. Of course, more work is needed to determine if this generalizes to other design matrices (e.g., fewer and smaller groups) as well as other parameter spaces (e.g., more reliable random effects).

Centering Variables in the Social Accuracy Model

Centering variables in multilevel models can have profound effects on the interpretation of coefficients, particularly when there are cross-level interactions (see Kreft, de Leeuw, & Aiken, 1995, for an introduction to the effects of centering in multilevel models). Grand mean centering the target validity measure and the normative profile by subtracting the mean across both target validity measures and targets has no impact on the estimates of distinctive or

TABLE 3
Mean Monte Carlo Simulation of Parameter Estimates Based on Round-Robin D Using Standard Multilevel Software as a Function of Correlated Perceiver-Target Random Effects

		Perceiver-Target Random Effects Correlation (ρ_{PT})		
Model Parameter	*Parameter*	*0*	*.30*	*.50*
Random Effects (*SD*)				
Perceiver				
Intercept	1.41	1.41	1.42	1.42
Validity measure	0.10	0.10	0.10	0.10
Target				
Intercept	1.41	1.41	1.42	1.43
Validity measure	0.22	0.22	0.22	0.22
Residual *SD*	1.70	1.70	1.70	1.70
Fixed Effects				
Intercept	0.00	−0.00	0.00	0.00
Validity measure (VM)	0.20	0.20	0.20	0.20
Moderator	0.00	0.00	−0.00	0.00
VM × Moderator	0.10	0.10	0.10	0.10

Note. The intercept correlated .30 with the validity measure for both Perceiver and Target random effects in the population. A total of 3,000 simulations were conducted for each of the three different perceiver-target random effect correlations. Note that standard errors for the fixed effects did not vary as a function of the perceiver-target random effects correlation.

normative accuracy in Equations (2.1–2). However, estimates of the intercept will change in this case and provide a component analogous to Cronbach's (1955) elevation accuracy. This component is rarely of interest in substantive research. Grand mean centering is often recommended simply to reduce the computational complexity of the estimation process; the resulting model is deterministically equivalent to the uncentered model (Kreft et al., 1995; see also Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004, for similar examples).

A more substantively useful form of centering for the SAM is to center the target validity measures *within each item.* For instance, define $\tilde{T}Self_{jk}$ as the centered target self-report within each item (i.e., $\tilde{T}Self_{jk} = TSelf_{jk} - Mean_k$). Then Equation (2.1) would be estimated as $Y_{ijk} = \beta_{0ij} + \beta_{1ij}\tilde{T}Self_{jk} + \beta_{2ij}Mean_k + \varepsilon_{ijk}$. This centering (see Furr, 2008) creates orthogonal predictors in Equation (2.1) as well as ensuring that different targets have $\tilde{T}Self_{jk}$ response profiles that are uncorrelated on average across targets (see Biesanz & West, 2000, p. 432, for an empirical demonstration). For basic SAM analyses such as Equation (2.1–2) with no moderators, uncentered predictor values ($TSelf_{jk}$)

present an analysis parallel to the aggregated model (see Kreft & de Leeuw, 1998)—the estimates of distinctive accuracy are the same in the two models; the estimates of normative accuracy differ between the two models by the amount of the fixed effect for distinctive accuracy. In general, the use of centered validity measures ($\tilde{T}Self_{jk}$) will produce estimates that are easier to interpret, particularly if moderators are introduced that are related to distinctive accuracy.

SUMMARY AND CONCLUSIONS

Why examine a complicated analytical model such as the SAM? Why not simply examine correlations such as self-other agreement separately for each trait as much previous research has done? Indeed, as noted earlier, within the SAM the estimate of distinctive accuracy may be interpreted as the average level of distinctive accuracy across targets where this average is computed across the different traits. Thus the traditional approach examining correlations separately for each trait provides a correspondence to part of the SAM while retaining a certain analytical and conceptual simplicity. However, there are number of strong reasons for examining the full SAM as opposed to piecemeal analyses for each trait.

First, the SAM, by retaining all of the information within a single analysis, provides substantially more statistical power than an analysis conducted on a single trait. Recent meta-analytic reviews of the accuracy of interpersonal perception across diverse contexts have demonstrated that accuracy exists for each of the Big Five personality traits (e.g., Hall, Andrzejewski, Murphy, Schmid Mast, & Feinsten, 2008; Holleran & Mehl, 2009). Thus averaging across different personality attributes represents a meaningful quantity of accuracy for personality traits. However, if differences between personality traits are of interest, instead of examining each trait in isolation, this analysis can be integrated with SAM. For instance, suppose that ratings of the public observability of different personality traits were obtained. Instead of examining self-other agreement correlations separately for each personality trait and then examining if these correlations were higher for more publicly observable traits, the ratings of public observability could be introduced directly with SAM as a moderating variable (e.g., see Equation (2.3); Human & Biesanz, 2010). Second, correlational analyses based on each trait omit substantially useful information. Effects related to stereotype accuracy are contained within the mean level and essentially lost when accuracy is evaluated for different isolated attributes. Stereotype accuracy is a real and important form of accuracy (e.g., see Funder, 2001). Appropriately assessing and evaluating the normativeness of impressions viewed can illuminate important interpersonal consequences. For instance, Biesanz and Human (2010) demonstrated that perceivers motivated to form accurate impressions in a study similar in design to Study 1 were able to form more distinctively accurate impressions when compared with perceivers not given this explicit social goal. However, accuracy-motivated interviewers perceived targets less normatively, and, specifically, less positively. Without evaluating normative accuracy directly, this cost for forming accurate impressions would not be apparent. Interestingly, normative accuracy, from the perceiver's perspective, represents both an evaluative component coupled with generalized knowledge. Because in the present studies the average self-report response profile is highly desirable, higher levels of perceiver normative accuracy may reflect a positive evaluative tendency. At the same time, it may reflect better generalized knowledge—a keener understanding of the commonalities and generalities in behavior across different individuals.

Third, modeling individual differences in the components of accurate interpersonal perception is not feasible through traditional analyses. Although it is possible to model individual differences for perceivers for certain designs (e.g., see Carlson & Furr, 2009), the strong benefit of the SAM is that it allows detailed analysis of individual differences among targets. This is only feasible through modeling multiple different traits simultaneously as Cronbach (1955) originally proposed.[8] As evidenced in Figure 5, there are substantial individual differences for both perceivers and targets with the latter, in particular, being relatively unstudied.

These reasons present a cogent argument for examining individual differences in accurate interpersonal perception through the SAM. However, it is ultimately the actual utility of a model that determines its eventual success, not its theoretical elegance. Recent applications of SAM across diverse contexts suggest that not only is SAM an elegant and concise analytical approach, its application can reveal interesting and nuanced findings. For instance, as discussed previously, Biesanz and Human (2010) randomly assigned perceivers to either the social goal of forming accurate impressions or no specific social goal. Accuracy-motivated perceivers achieved higher levels of distinctive self-other agreement but lower levels of normative agreement compared with perceivers not given an explicit impression formation goal. Those perceivers motivated to form accurate impressions did indeed become more accurate but at the cost of seeing others less normatively and, in particular, less positively. Chan et al. (2010) examined gender differences in accurate interpersonal perception for broad personality traits. Combining different studies to achieve a substantial sample size of 896 perceivers revealed no discernible gender difference in distinctive accuracy but

[8]Cronbach (1958) later focused on analyzing components of accuracy separately for each trait; although useful for certain questions and designs, a narrow trait-focused approach precludes examination of target effects. When the analysis is restricted to a single trait, there exists but a single validity measure value for the target on that trait. It is not possible to estimate relationships for the target based on a single predictor value.

a sizable difference in normative accuracy, with women's perceptions of others being more normative than those from men. Whether this result is due to differences in evaluative tendencies or generalized knowledge will be the focus of future research. Human and Biesanz (2010; see also Human, 2009) documented how diverse measures of adjustment are associated with perceptive accuracy and extend the SAM to examine assumed similarity. Additional findings include relationships between perceptions of target attractiveness and both distinctive and normative accuracy (Lorenzo, Biesanz, & Human, in press). Finally, perceiver assessments of both the confidence and the validity of their impressions were correlated with the actual distinctive accuracy of perceiver impressions (Biesanz et al., 2010).

In sum, the SAM offers a flexible data analytic model with parameters that represent constructs that are often of substantive interest—individual differences in components of accurate interpersonal perception for both perceivers and targets. Preliminary evidence suggests as well that standard multilevel modeling software is sufficient for estimating this model for round-robin data. Recent research demonstrates the practical utility of the model and how it can reveal relationships and effects obscured by traditional data analytic approaches. At long last, we can provide strong answers to the classic questions of who is the good judge and the good target. Beyond these classic questions, the flexibility of the SAM allows researchers to answer entirely new questions that have not been previously posed.

ACKNOWLEDGMENTS

Parts of this research were presented at the 2007 and 2009 Society of Multivariate Experimental Psychological Association's annual meetings (Biesanz, 2007, 2009). This research was partially supported by Social Sciences and Humanities Research Council of Canada Grant SSHRC 410-2005-2287 and 410-2008-2643 to Jeremy C. Biesanz. Thanks to David A. Kenny and R. Michael Furr for helpful comments on an earlier draft as well as Stephen G. West, Lauren J. Human, Katherine H. Rogers, Meanne Chan, Juliet Sarracino, Genevieve L. Lorenzo, Kate L. Parisotto, Caitlyn Moller, Annie-Claude Paquin, and the members of the Social Accuracy Laboratory for their assistance and insightful comments.

REFERENCES

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55,* 387–395.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64,* 431–441.

Bates, D., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes (*R* package version 0.9975-12) [Computer software]. Retrieved January 15, 2009, from http://cran.r-project.org

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11,* 142–163.

Biesanz, J. C. (2007, October). *Explorations in componential analyses of interpersonal perception data.* Cattell Award address presented at the Society of Multivariate Experimental Psychology, Chapel Hill, NC.

Biesanz, J. C. (2009, September). *The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy.* Paper presented at the Society of Multivariate Experimental Psychology, Salishan, OR.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9,* 30–52.

Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions: Accuracy motivated perceivers see the personality of others more distinctively but less normatively. *Psychological Science, 21,* 589–594.

Biesanz, J. C., Paquin, A., Sarracino, J., Human, L. J., Chan, M., & Gillis, R. (2010). *Are we aware of when our first impressions are accurate?* Manuscript in preparation.

Biesanz, J. C., & West, S. G. (2000). Personality coherence: Moderating self-other profile agreement and profile consensus. *Journal of Personality and Social Psychology, 79,* 425–437.

Biesanz, J. C., West, S. G., & Graziano, W. G. (1998). Moderators of self-other agreement: Reconsidering temporal stability in personality. *Journal of Personality and Social Psychology, 75,* 467–477.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology, 92,* 119–135.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology, 34,* 164–181.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62,* 645–657.

Borkenau, P., & Zaltauskas, K. (2009). Effects of self-enhancement on agreement on personality profiles. *European Journal of Personality, 23,* 107–123.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modeling, 1,* 103–124.

Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy: People understand the different impressions they make. *Psychological Science, 20,* 1033–1039.

Chan, M., Rogers, K., Parisotto, K. L., & Biesanz, J. C. (2010). *Forming first impressions: The role of gender and normative accuracy in personality perception.* Manuscript submitted for publication.

Colvin, C. R. (1993a). Childhood antecedents of young-adult judgability. *Journal of Personality, 61,* 611–635.

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

Colvin, C. R., & Bundick, M. J. (2001). In search of the good judge of personality: Some methodological and theoretical concerns. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 47–65). Mahwah, NJ: Erlbaum.

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

Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353–379). Stanford, CA: Stanford University Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 144–168). New York: Guilford.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science, 2,* 31–39.

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

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic.

Funder, D. C. (2001). Three trends in current research on person perception: Positivity, realism, and sophistication. In J. Hall & F. Bernieri (Eds.), *Interpersonal sensitivity: Theory, measurement and applications* (pp. 319–331). Mahwah, NJ: Erlbaum.

Funder, D. C., & West, S. G. (Eds.). (December, 1993). Viewpoints on personality: Consensus, self-other agreement and accuracy in judgments of personality [Special Issue]. *Journal of Personality, 61*(4).

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality, 76,* 1267–1316.

Furr, R. M. (2009). Profile analysis in person-situation integration. *Journal of Research in Personality, 43,* 196–207.

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62,* 411–422.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 89–150). New York: McGraw-Hill.

Gill, P. S., & Swartz, T. B. (2001). Statistical analyses for round robin interaction data. *The Canadian Journal of Statistics, 29,* 321–331.

Gill, P. S., & Swartz, T. B. (2007). Bayesian analysis of dyadic data. *American Journal of Mathematical and Management Sciences, 27,* 73–92.

Hall, J. A. (2001). The PONS test and the psychometric approach to measuring interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity*: *Theory and measurement* (pp. 143–160). Mahwah, NJ: Erlbaum.

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

Hall, J. A., & Bernieri, F. J. (Eds.). (2001). *Interpersonal sensitivity: Theory and measurement.* Mahwah, NJ: Erlbaum.

Holleran, S. E., & Mehl, M. R. (2009, July). *The accuracy of personality judgments at zero-acquaintance: A meta-analysis using realistic, everyday environments.* Poster presented at the Association for Research in Personality, Evanston, IL.

Human, L. J. (2009). *The role of adjustment in accurate interpersonal impressions.* Unpublished master's thesis, University of British Columbia, Vancouver, British Columbia, Canada.

Human, L. J., & Biesanz, J. C. (2010). *The role of adjustment in expressive accuracy: The good target or the good self judge?* Unpublished manuscript.
Human, L. J., & Biesanz, J. C. (in press). Through the looking glass clearly: The role of adjustment in accuracy and assumed similarity in first impressions. *Journal of Personality and Social Psychology.*
Human, L. J., Biesanz, J. C., Parisotto, K. L., & Dunn, E. W. (2010). *Your best self reveals your true self: Trying to make a good impression results in more accurate impression.* Manuscript submitted for publication.
Ickes, W. (Ed.). (1997). *Empathic accuracy*. New York: Guilford.
John, O. P., & Srivastava, S. (1999). The Big Five taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford.
Kenny, D. A. (1994). Interpersonal perception: A social relations analysis. New York: Guilford.
Kenny, D. A. (1998). SOREMO: A FORTRAN program for the analysis of round robin data structures [Computer software]. Storrs: University of Connecticut.
Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102,* 390–402.
Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8,* 115–128.
Kenny, D. A., & LaVoie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142–182). San Diego, CA: Academic.
Kenny, D. A., West, T. V., Malloy, T. E., & Albright, L. (2006). Componential analysis of interpersonal perception data. *Personality and Social Psychology Review, 10,* 282–294.
Kenny, D. A., & Winquist, L. A. (2001). The measurement of interpersonal sensitivity: Consideration of design, components, and unit of analysis. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 265–302). Mahwah, NJ: Erlbaum.
Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43,* 23–34.
Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30,* 1–21.
Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (in press). What is beautiful is good and more accurately understood: Physical attractiveness and accuracy in first impressions of personality. *Psychological Science.*
MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1,* 173–181.
Magnusson, D., & Törestad, B. (1993). A holistic view of personality: A model revisited. *Annual Review of Psychology, 44,* 427–452.
McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90,* 215–238.
McDonald, R. P. (1981). Constrained least squares estimates of oblique common factors. *Psychometrika, 46,* 337–341.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4,* 681–691.
Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211–238). New York: Guilford.
Paulhus, D. L. (2009). [Social desirability of personality measures]. Unpublished raw data.
R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software]. Vienna: *R* Foundation for Statistical Computing. Retrieved January 15, 2009, from http://www.R-project.org

Shen, W., & Louis, T. A. (1999). Empirical Bayes estimation via the smoothing by roughening approach. *Journal of Computational and Graphical Statistics, 8,* 800–823.

Skowronski, J., & Ambady, N. (Eds.). (2008). *First impressions*. New York: Guilford.

Steiger, J. H. (2006a, June). *Things we could have known: Some thoughts on seeing the future and avoiding regret in data analysis and model selection*. Keynote address, International Meeting of the Psychometric Society, Montreal, QB, CA.

Steiger, J. H. (2006b, October). *Things we could have known: Some thoughts on seeing the future and rediscovering the past in data analysis and model selection.* Presidential address, Society of Multivariate Experimental Psychology, Lawrence, KS.

Swann, W. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91,* 457–477.

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round-robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology, 37,* 1742–1757.

Wong, G. Y. (1982). Round robin analysis of variance via maximum likelihood. *Journal of the American Statistical Association, 77,* 714–724.

Wood, D., Gosling, S. D., & Potter, J. (2007). Normality evaluations and their relation to personality traits and well-being. *Journal of Personality and Social Psychology, 93,* 861–879.