# What Do You Learn About Someone Over Time? The Relationship Between Length of Acquaintance and Consensus and Self–Other Agreement in Judgments of Personality

Jeremy C. Biesanz
University of British Columbia

Stephen G. West and Allison Millevoi
Arizona State University

Theory and research examining length of acquaintance and consensus among personality judgments have predominantly examined each dimension of personality separately. In L. J. Cronbach's (1955) terminology, this trait-centered approach combines consensus on elevation, differential elevation, and differential accuracy in personality judgments. The current article extends D. A. Kenny's (1991, 1994) weighted average model (WAM)—a theoretical model of the factors that influence agreement among personality judgments—to separate out two of Cronbach's components of consensus: stereotype accuracy and differential accuracy. Consistent with the predictions based on the WAM, as length of acquaintance increased, self–other agreement and consensus differential accuracy increased, stereotype accuracy decreased, and trait-level or raw profile correlations generally remained unchanged. Discussion focuses on the conditions under which a relationship between length of acquaintance and consensus and self–other agreement among personality evaluations emerges and how impressions change over time.

*Keywords:* personality, consensus, acquaintance, accuracy, stereotype accuracy

Upon meeting someone for the first time, it is natural to form an initial impression of that person. As an acquaintance develops, our intuitive sense is that we come to know the person better. Evidence consistent with this intuitive sense is provided by a large body of research showing that acquaintances are better able to predict behavior and to agree with self-reports than are strangers who observe only several minutes of behavior (e.g., Colvin & Funder, 1991; Funder & Colvin, 1988; Funder, Kolar, & Blackman, 1995; Jackson, Neill, & Bevan, 1973; Norman & Goldberg, 1966; Paulhus & Bruce, 1992; Paulhus & Reynolds, 1995; Watson, 1989; Watson & Clark, 1991; see Funder & West, 1993, for a detailed discussion of issues in consensus, self–other agreement, and accuracy).

After those few initial encounters, is there any evidence consistent with this intuitive sense that we come to know the person better with further contact? For example, do our judgments of personality traits show more agreement with those of the target and with other knowledgeable informants after two years as compared with just several weeks? It is interesting to note that longitudinal research has generally revealed little support for enhanced consensus among acquaintances

over time (e.g., Kenny, Albright, Malloy, & Kashy, 1994; Park, Kraus, & Ryan, 1997; Paulhus & Reynolds, 1995). Yet other research has found self–other agreement increasing with levels of acquaintance (e.g., Kurtz & Sherker, 2003; Paulhus & Bruce, 1992; Paunonen, 1989; Watson & Clark, 1991; Watson, Hubbard, & Wiese, 2000). Given the centrality of personality judgments in psychological research, understanding when such judgments are reliable, informative, and accurate remains a critical question.

The present research examines the impact of increased length of acquaintance on consensus in judgments of personality among existing acquaintances and agreement with self-reports. First, we briefly review the literature on length of acquaintance and consensus and self–other agreement. Next we consider two theoretical models of the development of consensus in personality judgments, Funder's (1995, 1999) realistic accuracy model (RAM) and Kenny's (1991, 1994) weighted average model (WAM), and examine the WAM's predictions for raw profile consensus and for Cronbach's (1955) consensus components of stereotype accuracy and differential accuracy. We then argue, building on these models, that increased acquaintance is likely to be related to consensus and self–other agreement among personality judgments when (a) one, but not both, of the judgments is already highly reliable (Kenny, 2004); (b) judgments are based on observations across numerous different situations; and/or (c) consensus and self–other agreement are examined through differential accuracy. Finally, we present two studies examining the relationship between length of acquaintance and different components of consensus and self–other agreement.

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

## Length of Acquaintance and Agreement

Empirical research on the relationship between the length of acquaintance and the level of consensus has yielded a complex

pattern of results. Although knowledgeable informants reach higher levels of consensus than do strangers in cross-sectional studies (e.g., Albright et al., 1988; Ambady & Rosenthal, 1992; Ambady, Hallahan, & Rosenthal, 1995; Borkenau & Liebler, 1992; Levesque & Kenny, 1993; Norman & Goldberg, 1966; Zebrowitz & Collins, 1997), Kenny et al.'s (1994) review of longitudinal studies found no evidence for a relationship between length of acquaintance and consensus among existing acquaintances. More recent longitudinal studies are consistent with this lack of relationship (e.g., Park et al., 1997; Paulhus & Reynolds, 1995).

In contrast to studies that examine consensus among acquaintances, *self–other* agreement does appear to increase with length of acquaintance (e.g., Funder & Dobroth, 1987; Park & Judd, 1989; Paulhus & Reynolds, 1995; Paunonen, 1989; Watson & Clark, 1991; Watson et al., 2000). Note that this effect has generally been observed with cross-sectional designs. Longitudinal studies typically enable stronger inferences than do cross-sectional studies in which acquaintances are nested within participants, and thus length of acquaintance is confounded with participant (cf. Kenny & Albright, 1987). Although a nested design potentially raises the possibility that factors other than length of acquaintance may account for enhanced levels of self–other agreement, alternative explanations (e.g., greater similarity among self–acquaintance pairs) do not appear to be viable (see, e.g., Funder, Kolar, & Blackman, 1995). This conclusion is further supported by longitudinal research demonstrating enhanced self–other agreement over time. For example, Paulhus and Bruce (1992) examined agreement within discussion groups over the course of 7 weeks. Although consensus among acquaintances did not increase significantly, self–other agreement did. This longitudinal relationship for self–other agreement has since been replicated (e.g., Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Kurtz & Sherker, 2003; however, see Park et al., 1997, for a failure to replicate). To understand these empirical results, and to examine the conditions under which length of acquaintance might be related to consensus and self–other agreement, we consider two theoretical perspectives on personality judgments: the realistic accuracy and the weighted average models.

### Realistic Accuracy Model

Funder's (1995, 1999) RAM outlines the process of how an accurate personality judgment is made. Starting with three core premises—(a) that personality traits exist, (b) that people sometimes make personality judgments of others, and (c) that judgments are at least sometimes accurate—RAM describes four components in the process underlying the formation of an accurate judgment. The person being judged must behave in a manner relevant to the trait being judged in a way that is also available to the perceiver. In turn, this behavior must be detected and then utilized in forming a personality judgment. These four components of an accurate judgment (*relevance*, *availability*, *detection*, and *utilization*) are all essential. As predicted by RAM, factors that impact relevance, availability, detection, and utilization have all been shown to influence levels of agreement (see Funder, 1995, 1999, for reviews).

RAM focuses on the formation of accurate personality judgments. How does this help us to understand the relationship between length of acquaintance and consensus? As Funder (1999, pp.

157–158) noted, consensus can be achieved in the absence of accuracy, yet a sufficiently high level of accuracy demands consensus among observers. If length of acquaintance improves accuracy, then a relationship to consensus will eventually emerge although the RAM model is not sufficiently explicit to detail the exact nature of this relationship.

### Weighted Average Model

Kenny (1991, 1994) originally proposed the WAM, an impression formation model that can be used to examine how different factors impact the level of consensus among observers. Kenny (2004) has recently reparameterized this into a personality, error, residual, stereotype, opinion, and norm (PERSON) model for the social relations model; nonetheless, both the original and revised forms of the model result in the same predicted relationships. On the basis of Anderson's (1981) weighted averaging model of impression formation, WAM is a mathematical model that provides precise predictions for when and how length of acquaintance is related to consensus. The WAM assumes that acquaintances observe behavioral acts ($A$), assign a scale value to each act ($s$), and form an impression ($I$) that is simply a weighted average of these scale values across observed acts combined with weighted stereotype and unique impression components. For two acquaintances who observe a series of targets engaging in a series of acts, the correlation between the acquaintances' impressions may be expressed as follows,[1] (Kenny, 1994, p. 248):

$$\rho = \frac{w^2 \rho_4 + 2wn\rho_6 + qn\rho_2 + (n^2 - qn)\rho_3}{(n^2 - n)\rho_1 + 2wn\rho_5 + n + k^2 + w^2}. \quad (1)$$

Under this model, consensus between acquaintances across targets on a single variable is comprised of the following factors.

### Acquaintance (n)

*Acquaintance* is simply the number of observed acts. To simplify the formula, Equation 1 presumes that this number is the same for each observer.

### Overlap (q)

*Overlap* is the proportion of acts that are observed in common.

### Stereotypes (w) and Extraneous Information (k)

Both $w$ and $k$ index the weight of stereotypical and extraneous (unique) information, respectively, on impressions.

### Consistency Within an Observer ($\rho_1$)

*Consistency within an observer* refers to the stability of scale weights within a target across different acts for a single observer. That is the correlation of scale weights for a given target across acts from the perspective of a single observer. As is discussed in more detail shortly, this is a function of a target's behavioral

---

[1] For ease of presentation, we make the simplifying assumption that there is no communication among acquaintances. This assumption has no material effect on the conclusions that are reached.

consistency, the range of observed acts, and the perceptual processes of the observer.

## Shared Meaning Systems ($\rho_2$)

*Shared meaning* is the correlation (agreement) between two acquaintance's scale weights when observing the same act.

## Consistency Between Observers ($\rho_3$)

*Consistency between observers* refers to the correlation between the scale weights of different acquaintances observing different acts. As with consistency within an observer, this is a function of a target's behavioral consistency, the range of observed acts, as well as shared meaning between acquaintances. In Kenny's (1991) original formulation, $\rho_3$ was constrained to equal to $\rho_1 \times \rho_2$, but consistent with Kenny (1994) we do not impose this constraint.

The last three terms are related to the impact of stereotypes on impressions. Given stereotypical information due to, for example, physical appearance or how people in general act, the stereotypical information influences consensus as a function of agreement between judges on the stereotype ($\rho_4$), the consistency within an observer between the stereotype and an act ($\rho_5$), and the consistency between one observer's evaluation and the other observer's stereotype ($\rho_6$). Components of WAM have been demonstrated to influence consensus as predicted (e.g., Chaplin & Panter, 1993; Malloy, Agatstein, Yarlas, & Albright, 1997; Park, DeKay, & Kraus, 1994; Story, 2003). More specifically, with respect to length of acquaintance, WAM predicts that consensus among acquaintances emerges quickly and then levels off (asymptotes), and agreement with a knowledgeable other (e.g., self–other agreement) increases and eventually asymptotes with most of the increase occurring early. These predictions are consistent with previous empirical research.

### Reexamining Length of Acquaintance and Consensus

Examining the WAM carefully reveals an interesting relationship between the length of acquaintance and trait-level consensus. If consistency within and between observers ($\rho_1$ and $\rho_3$, respectively) are both positive and greater than, for example, $\rho_1 = \rho_3 = .10$, then consensus among acquaintances may asymptote very quickly and further increases in the length of acquaintance have essentially no impact on consensus. Of importance, as consistency within an observer ($\rho_1$) decreases, then agreement among observers will manifest itself more slowly (see Kenny, 1991, Figure 3). Thus conditions that paradoxically decrease $\rho_1$, the consistency within observers across acts, are precisely those conditions under which a relationship between length of acquaintance and agreement is more likely to emerge—particularly when $\rho_1$ is already low.

We consider three specific circumstances in which length of acquaintance is more likely to be related to consensus than would be expected from general summaries of the WAM formulation. First, we consider the impact of observing behavior under a variety of different situations. Second, we examine the predictive accuracy of an acquaintance's impressions. For example, how does the correspondence between an acquaintance's impressions and criteria such as behavioral measures or self-reports change over time?

Third, we reformulate the WAM to examine profile consensus across different traits and Cronbach's (1955) components of consensus of stereotype accuracy and differential accuracy. The WAM, as expressed in Equation 1, models consensus among different acquaintances for a single trait, which, following Cronbach (1955), combines multiple different components of consensus (see also Kenny & Winquist, 2001). By incorporating insights from the RAM and Cronbach's componential analysis of accuracy, new predictions emerge from the WAM as to how consensus changes over time for traditional trait-level consensus, raw profile consensus, differential accuracy, and stereotype accuracy.

## Natural Dyads Versus Assessment Groups

One of the striking differences between previous cross-sectional research and longitudinal research is the variety of different situations that acquaintances witness behavior. Cross-sectional research has used naturally occurring dyads in which observations are made across many different situations. Long-term acquaintances may interact in home, school, recreational, religious, leisure, and many other settings. In contrast, the longitudinal research to date has predominantly used groups in structured situations, such as study groups and experimental laboratory groups, for convenience in data collection. These groups severely constrain the set of situations under which behavior can be observed. When behavior is observed in similar situations over time, impression scale weights across these different situations will correlate highly (e.g., see Funder & Colvin, 1991), leading to high levels of $\rho_1$ and consequently the inability to detect a relationship between length of acquaintance and agreement. In contrast, if behavior is observed across many different situations, there will be substantially less consistency in the impressions across different observed acts. To the extent that behavior is observed in many different situations, such as with naturally occurring dyads, we would expect agreement to increase with length of acquaintance.

## Acquaintance-Criterion Agreement

The WAM focuses on consensus among acquaintances, each of whom observes a series of acts from a common target. As the acquaintance develops, each observer's impression becomes more reliable, and agreement is examined among increasingly reliable judgments. Reliability is defined here in comparison to the theoretical impression that would be formed by observing the hypothetical universe of acts from the common target (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). In contrast, self–other and acquaintance-criterion agreement present two sharp differences from this account. First, only the acquaintance's impressions become appreciably more reliable over time. An elegant example of this is provided by Borkenau, Mauer, Riemann, Spinath, and Angleitner (2004). Targets were videotaped performing 15 different behavioral tasks, and each video clip was assessed by a separate judge. As the number of witnessed acts increased—here aggregated across different judges who each observed only one act—the relationship between self-reports, peer reports, and the experimenter and confederate in the study increased as well.

Second, from the perspective of the WAM, self-reports are inherently more reliable as they are based on many more (self) observations than are reports by acquaintances (e.g., Epstein,

1983; Funder & Colvin, 1997). As a consequence, we should expect higher levels of self–other agreement than consensus among acquaintances, holding the length of acquaintance constant. We would also expect higher eventual (asymptotic) levels of agreement.[2] Higher asymptotic levels of self–other agreement presents a greater opportunity to detect the emergence of such agreement as compared with the lower asymptotic levels theoretically achieved by consensus between acquaintances.

## Length of Acquaintance and Cronbach's Components of Consensus

In a highly influential article, Cronbach (1955) argued for partitioning the correspondence between a judgment and a criterion—such as self and acquaintance ratings across multiple traits—into four components: *elevation accuracy*, *differential elevation accuracy*, *stereotype accuracy*, and *differential accuracy*. *Elevation accuracy* is the correspondence between the grand mean of the judgments and the grand mean of the criterion (across multiple targets). *Differential elevation accuracy* is the correspondence between the mean judgment and the mean criterion for each target after removing each respective grand mean. *Stereotype accuracy* is the correspondence between the judgments across traits and the "generalized other" or the average person. Stereotype accuracy can be assessed by computing the profile correlation between a judges' set of ratings across different traits and the profile of the mean ratings across persons. *Differential accuracy* is the relationship between the unique component of judgments after removing the "generalized other" and the unique component of the criteria. Differential accuracy thus removes the correspondence attributable to stereotype accuracy. We now consider these four components in light of WAM.

As currently formulated, WAM is a variable-centered model. That is, consensus and self–other agreement are calculated separately for each trait across persons. For example, does increased length of acquaintance lead to enhanced consensus on extraversion *across* individuals? Virtually all research examining length of acquaintance and consensus and self–other agreement has been conducted by using such analyses that focus on a specific trait. Such trait-level analyses, to use Cronbach's terminology, combine elevation accuracy, differential elevation accuracy, and differential accuracy. Note that disentangling elevation accuracy and differential elevation accuracy from differential accuracy requires research designs in which acquaintances rate multiple different targets (see Kenny & Winquist, 2001, for a discussion of variable versus person-centered analyses, Cronbach's componential approach, and design considerations).

Only a small number of studies have examined other forms of consensus. For example, profile consensus across traits is greater for acquaintances than among strangers (Funder, Kolar, & Blackman, 1995). Self–other profile agreement increases with exposure, whereas profile consensus among observers asymptotes very quickly—at 30 min for observers watching videotapes in a laboratory (Blackman & Funder, 1998). Both of these studies examined raw profile agreement, which combines stereotype accuracy with differential accuracy. Bernieri et al. (1994) found that length of cohabitation among college roommates was related to raw profile self–other agreement and differential accuracy but not stereotype accuracy. This limited research on the influence of

length of acquaintance with profiles so far has been consistent with the results from similar studies focused on trait-level consensus and self–other agreement. However, should we expect the exact same relationships and effects of length of acquaintance on consensus for trait-level analyses and for raw profile consensus, stereotype agreement, and differential accuracy? We now reformulate the WAM to examine the impact of length of acquaintance for these different forms of consensus and accuracy.

Consider the traditional trait-based formulation of the WAM illustrated in Figure 1A. Observers witness a target's acts, only some of which are seen in common, and assign scale weights to these acts for a particular trait dimension (e.g., extraversion). Consensus for two judges is examined across a series of targets. However, the same act can have trait implications for more than one trait dimension. For example, consider two acts: (a) extensive preparation for a job interview and (b) attendance at the gala opening of an avant-garde art gallery. The first act has implications for both conscientiousness and neuroticism (anxiety) and the second act for both extraversion and openness to experience. Thus, each act can have scale weights for multiple trait dimensions. Figure 1B illustrates such a WAM across different trait dimensions. Raw profile consensus and self–other agreement can then be calculated between two observers for the same target across a set of trait dimensions.

It is interesting to note that Equation 1, developed to describe the parameters influencing consensus between judges for a single trait dimension, also serves to describe the parameters affecting raw profile agreement between two judges. However, the interpretation of three critical parameters ($\rho_1$, $\rho_2$, and $\rho_3$) changes in important ways. *Consistency within an observer* ($\rho_1$) now refers to the profile correlation of scale weights across the set of traits assigned by the observer for different acts. That is, for two different acts, $\rho_1$ measures the relationship between the assigned scale weights across different trait dimensions. *Shared meaning* ($\rho_2$) is the profile correlation of the scale weights for two different observers who witness the same act. *Consistency between observers* ($\rho_3$) now refers to the profile correlation of scale weights between two observers who witness different acts.

Observed levels of raw profile consensus between two observers may be due in part to stereotype accuracy—people are in fact similar to each other, on average, with respect to their personality profile (e.g., McCrae et al., 2005). Indeed, Blackman and Funder (1998) noted an average profile correlation between two observers rating different individuals of .16. Explicitly separating out stereo-

---

[2] Consensus among acquaintances (assuming no overlap) asymptotes to approximately $\frac{\rho_3}{\rho_1}$. If $\rho_3 = \rho_2 \times \rho_1$, as Kenny (1991) assumed, this reduces to shared meaning for the same single act. In contrast, self–acquaintance agreement will asymptote to approximately $\rho_{sc} \sqrt{\rho_{cc}}$, where $\rho_{sc}$ is the correlation between the acquaintance's scale weights for a single act and the criterion measured without error, and $\rho_{cc}$ is the reliability of the criterion. If the criterion (e.g., self-report) is reliable and $\rho_{sc}$ is greater than shared meaning ($\rho_2$), then self–acquaintance agreement will asymptote to higher levels than acquaintance consensus. Note also that this excludes the impact of agreement due to stereotypes and appearance (see Kenny, 1994, 2004), which may have a profound impact on consensus.
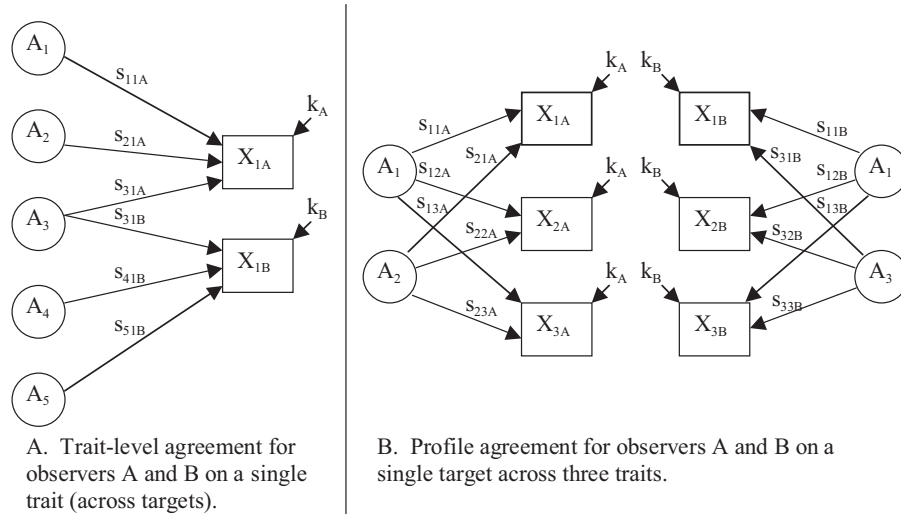
A. Trait-level agreement for observers A and B on a single trait (across targets).

B. Profile agreement for observers A and B on a single target across three traits.

*Figure 1.* Weighted average model for both trait-level consensus and profile consensus between observers A and B. Impressions ($X_{jl}$) are an aggregation of the scale values ($s_{ijl}$) assigned to behavioral acts ($A_i$) and unique impressions ($k_l$) where $i$ refers to the behavioral act; $j$, to the specific trait under consideration; and $l$, to the observer. For ease of representation, we have treated behavioral acts as variables and scale values as paths, conventions which differ from those traditionally used in the WAM (see Appendix A).

type accuracy and differential accuracy from raw profile consensus reveals an interesting set of predictions as illustrated in Figure 2.[3]

In this example, stereotype accuracy is initially high and increases slightly at first, but it diminishes over time as impressions become more individuated. Raw profile consensus exists at initial acquaintance because of stereotype accuracy, but it does not increase much over time as consensus quickly approaches the asymptotic level. Differential accuracy, in contrast, slowly increases and asymptotically reaches the same level as raw profile consensus. Figure 2 provides a graphical illustration of Funder's (1999) argument regarding how consensus may remain unchanged as differential accuracy increases.

Length of acquaintance, according to the WAM, is more strongly related to differential accuracy than to raw profile consensus given strong levels of stereotype accuracy. We also predict that this relationship will be stronger for differential accuracy than for trait-level consensus, which combines differential accuracy with elevation and differential elevation accuracy. If we correlate scale weights across traits within the same observer, this will result in theoretically meaningful differences in the expected relationship between length of acquaintance and consensus for differential accuracy versus trait-level analyses. On the basis of implicit personality theory (cf. Borkenau, 1992), the notion that such associations exist within a perceiver seems likely given the moderate to strong intercorrelations among personality assessments from the same source that are not present across different reporting sources (see Biesanz & West, 2004). For example, the scale weight assigned to extraversion for the act of attendance of the art gallery gala will be correlated with the scale weight assigned to openness to experience. To the extent that scale weights for the same act are correlated across traits, $\rho_1$ will be lower for differential accuracy analyses than for trait-level analyses (see Appendix A for a more formal derivation of this argument). Consequently, we predict that differential accuracy will emerge more slowly than trait-level

consensus, making it more likely to reveal a relationship with length of acquaintance.

## Overview of Studies 1 and 2

Across two studies we examined the relationship between length of acquaintance and consensus between peers as well as between peers and parents and self–other agreement in which the other is a peer or a parent. The present studies expand on previous research by (a) extending the WAM to compare the results of trait-level consensus, raw profile consensus, differential accuracy, and stereotype accuracy, (b) considering multiple different forms of self–other agreement and consensus, and (c) examining the emergence of agreement over long periods of acquaintance. Study 1 uses a cross-sectional design in which each set of acquaintances is nested within each target person. For Study 2, target persons were specifically asked to select multiple shorter term and longer term acquaintances in a within-subject design. For both studies, we predicted that length of acquaintance would be (a) positively

---

[3] To derive these predictions, and following Cronbach (1955), we consider the stereotype to be the average (mean) profile that both observers share (i.e., $\rho_4 = 1$). The resulting equation for stereotype accuracy (SA) under WAM is $\rho_{SA} = \dfrac{w + n\rho_5}{\sqrt{k^2 + w^2 + n + n(n-1)\rho_1 + 2w\rho_5}}$. As Kenny (1994, p. 247) noted, this will asymptote to $\dfrac{\rho_5}{\sqrt{\rho_1}}$. Differential accuracy (DAr), computed by subtracting the average profile from each observer's set of ratings, results in the same relationship as Kenny (1991, Equation 2): $\rho_{DAr} = \dfrac{qn\rho_2 + (n^2 - qn)\rho_3}{(n^2 - n)\rho_1 + n + k^2}$. The parameters used to produce Figure 2 are $\rho_1 = .04$, $\rho_2 = .1$, $\rho_3 = .01$, $\rho_4 = 1$, $\rho_5$ and $\rho_6 = .05$, $q = 0$ (no overlap), $w = 5$, and $k = 10$. These values were chosen so that the asymptotic profile relationships would correspond with previous estimates (e.g., Biesanz & West, 2000; Blackman & Funder, 1998).
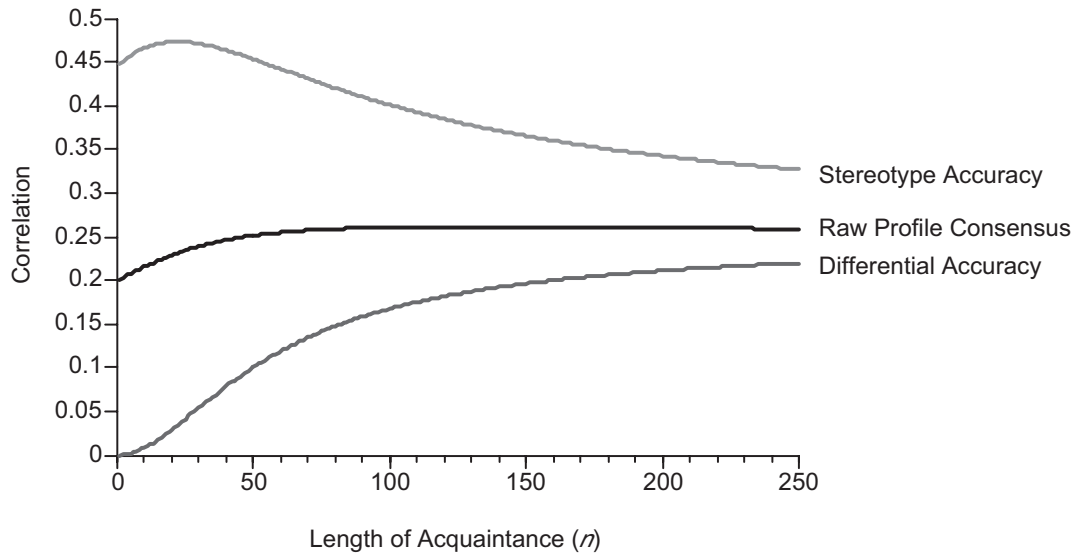
*Figure 2.* Ilustration of the theoretical relationship between length of acquaintance and raw profile consensus, stereotype accuracy, and differential accuracy.

related to differential accuracy, (b) not related to raw profile consensus and self–other agreement, and (c) negatively related to stereotype accuracy. We predicted that the strength of this relationship with differential accuracy would be strongest for self–acquaintance and parent–acquaintance agreement, cases in which the self- or parent informant would be expected to have a particularly reliable judgment of the target person.

## Study 1

### Method

#### Participants

Introductory psychology students ($N = 387$) were recruited to participate in return for partial fulfillment of their class requirements. A total of 339 participants completed the basic study requirements of attending three measurement sessions (226 women and 113 men; $M$ age = 19.48 years, $SD = 3.05$). Participants provided consent for obtaining a parental rating via mail and were encouraged to bring two acquaintances into the laboratory in exchange for additional credit toward fulfillment of their course requirements. Participants occasionally brought individuals whom they identified as relatives or romantic partners rather than acquaintances; the data from these individuals were excluded. A total of 266 (177 women and 89 men) participants had at least one acquaintance rating, and, of these, 193 (134 women and 59 men) had a parental rating with complete data. Average length of acquaintance was 21.91 months ($SD = 30.05$). Previous reports have appeared on the basis of this data set that address other questions (Biesanz & West, 2000 [Study 2], 2004).

#### Materials, Design, and Procedure

Participants rated themselves on 97 unipolar trait adjectives—20 for Agreeableness, 19 for Conscientiousness, 20 on Extraversion, 18 on Neuroticism, and 20 for Openness to Experience (Goldberg, 1992). Three trait adjectives proposed by Goldberg (1992)—*imperturbable*, *haphazard*, and *unexcitable*—were not included because they were unfamiliar or confusing to a large proportion of the respondents (cf. Biesanz & West, 2000). All

ratings were on a 9-point scale ranging from 0 (*extremely inaccurate*) to 8 (*extremely accurate*). Participant's self-rating instructions were modified from Goldberg (1992) to limit self-assessments of behavior to the previous week (see Biesanz, West, & Graziano, 1998, for the specific instructions). Participants completed the self-report inventory three times, at no less than 1-week intervals, in a lecture hall reserved for that purpose. Self-assessments were aggregated across the three assessments to yield a more trait-like measure, and a full table of the correlations among measures and internal reliabilities is presented in Biesanz and West (2004, Table 4).

Acquaintances and parents rated the participant on the same 97 unipolar trait adjectives by using the same 9-point rating scale and Goldberg's (1992) standard rating instructions, with the participant's name embedded within the instructions. Acquaintances indicated how long they had known the participant with the choices being 1–3 months, 3–6 months, 6–9 months, 9–12 months, 1–2 years, 2–4 years, 4–6 years, 6–10 years, or 10+ years. Length of acquaintance, measured in months, was determined by taking the midpoint of the interval range selected by the acquaintance, with 120 months being the maximum length coded.

#### Analytic Strategy

A subset of participants had two acquaintance ratings available. Although the choice of which of these acquaintance ratings to include in analyses with participants who had only one acquaintance rating is arbitrary, obtained results may vary slightly depending on the choice. Therefore, for participants with two acquaintance ratings, we randomly selected which acquaintance to include in the analyses. We repeated this randomization process 10 times, which resulted in 10 distinct data sets. Each data set was analyzed separately, and the results were then combined by using the procedures outlined in Rubin (1987) and Schafer and Graham (2002). This procedure allows the full use of all available data without limiting inferences to those participants with two acquaintance ratings.

### Results

Does knowing someone longer lead to enhanced consensus and agreement? To address this question, we first examine agreement

on mean trait levels and then examine the profile measures of raw profile consensus, stereotype accuracy, and differential accuracy.

## Length of Acquaintance and Trait-Level Consensus and Self–Other Agreement

Does knowing someone longer increase agreement on trait level? To examine whether length of acquaintance is related to self–acquaintance agreement and parent–acquaintance consensus on mean trait level, we conducted a series of multiple regression analyses separately for each of the Big Five personality factors, following the procedures outlined in Aiken and West (1991). Self- and parent trait-level ratings served as the criteria (for means and levels of agreement and consensus, see Biesanz & West, 2004). The predictors were acquaintance trait-level rating, length of acquaintance in months, and the acquaintance trait-level rating by length of acquaintance product term. The latter product term provides a test of whether length of acquaintance is associated with the amount of consensus in that a significant positive interaction term would indicate that consensus increases with length of acquaintance.

Across the Big Five personality traits, length of acquaintance did not significantly moderate self–acquaintance agreement (all $t$s < 1.20) or parent–acquaintance consensus (all $t$s < 1.65). Length of acquaintance marginally moderated parent–acquaintance consensus for Agreeableness, $t(189) = 1.65, p = .10$, such that parent–acquaintance consensus tended to increase with length of acquaintance.

## Length of Acquaintance and Raw Profile Self–Acquaintance Agreement and Parent–Acquaintance Consensus

Does knowing someone longer increase raw profile correlations? Self–acquaintance and parent–acquaintance consensus raw profile correlations were calculated for each participant across the full set of trait adjectives after reverse-coding items where appropriate. Note that all analyses were based on profile correlations that were first transformed to Fisher's variance-stabilizing $z_r$. On average, across participants, there were moderate to large levels of raw profile correspondence between participants and acquaintances (mean $r = .37, SD = .21, p < .001$) as well as between acquaintance and parents (mean $r = .36, SD = .20, p < .001$). However, length of acquaintance was not significantly associated with either self–acquaintance profile agreement, $r(264) = .10, ns$, or parent–acquaintance consensus, $r(191) = .13, ns$.

## Length of Acquaintance and Stereotype Accuracy

To compute stereotype accuracy, we correlated each acquaintance's report with the average parental report (i.e., the means of the parent-reported item levels). There were substantial levels of stereotype accuracy on average across acquaintances (mean $r = .46, SD = .25, p < .001$). However, contrary to prediction, length of acquaintance did not have a significant negative association with stereotype accuracy, $r(191) = -.06, ns$. Note that the obtained results were exactly equivalent when using the average self-report to compute stereotype accuracy.

## Length of Acquaintance and Differential Accuracy

To compute differential accuracy and to remove the inflating impact of stereotype accuracy, responses were standardized within reporting source before computing profile correlations. For example, parent responses to each item were standardized on the basis of the mean and standard deviation across all parental responses to that item (see Biesanz & West, 2000, for a more complete description of the procedure). Following these adjustments, on average there were small to moderate levels of differential accuracy between participants and acquaintances across participants (mean $r = .18, SD = .22, p < .001$) as well as between acquaintances and parents (mean $r = .13, SD = .21, p < .001$). As a check on these adjustment procedures, we computed adjusted profile correlations between random pairings of individuals and found that they did not differ significantly from zero (see Biesanz & West, 2000).

In contrast to the trait-level and raw profile correlational results, length of acquaintance was positively associated with both self–acquaintance differential accuracy, $r(264) = .18, p < .01$, and peer–parent differential accuracy, $r(191) = .16, p < .05$. On average, self–acquaintance and parent–acquaintance differential accuracy correlations increased by .09 and .07, respectively, for every 5-year increase in length of acquaintance. Given that the WAM model predicts that differential accuracy should approach an asymptote as length of acquaintance increases, we also examined the data by using a Lowess smoother, which provides a nonparametric estimate of the form of the relationship (see Cleveland, 1979; Cohen, Cohen, West, & Aiken, 2003; Cook & Weisberg, 1999). No evidence of nonlinearity was observed, and adding a quadratic component to length of acquaintance did not significantly improve the model for either self–acquaintance, $F(1, 260) = 0.21, ns$, or parent–acquaintance differential accuracy, $F(1, 190) = 0.17, ns$. Figure 3 presents the relationship between length of acquaintance and differential accuracy by using all available data and shows the linear increase in self–acquaintance and parent–acquaintance differential accuracy correlations.

## Discussion

Length of acquaintance was not significantly related to trait-level consensus or self–other agreement, raw profile consensus or self–other agreement, or stereotype accuracy. In contrast, consensus in differential accuracy between acquaintances and the parents of the target person and the agreement between the target person and the acquaintances did increase with time. After removing stereotype accuracy agreement and elevation effects, longer term acquaintances showed more agreement with participants as well as the participant's parents on the relative ordering of participants' personality attributes than did more recent acquaintances. In summary, knowing someone for a longer time did not lead to enhanced agreement with either that person or his or her parent on their level of extraversion. However, it did lead to enhanced agreement on whether they were more extraverted than they were conscientious. The use of differential accuracy correlations helped minimize the effects of potential differential scale usage by different informants that might partially obscure relationships at the trait-level where informants only report on one target.

Strong inferences based on these findings, however, are limited by several study design features. First, length of acquaintance
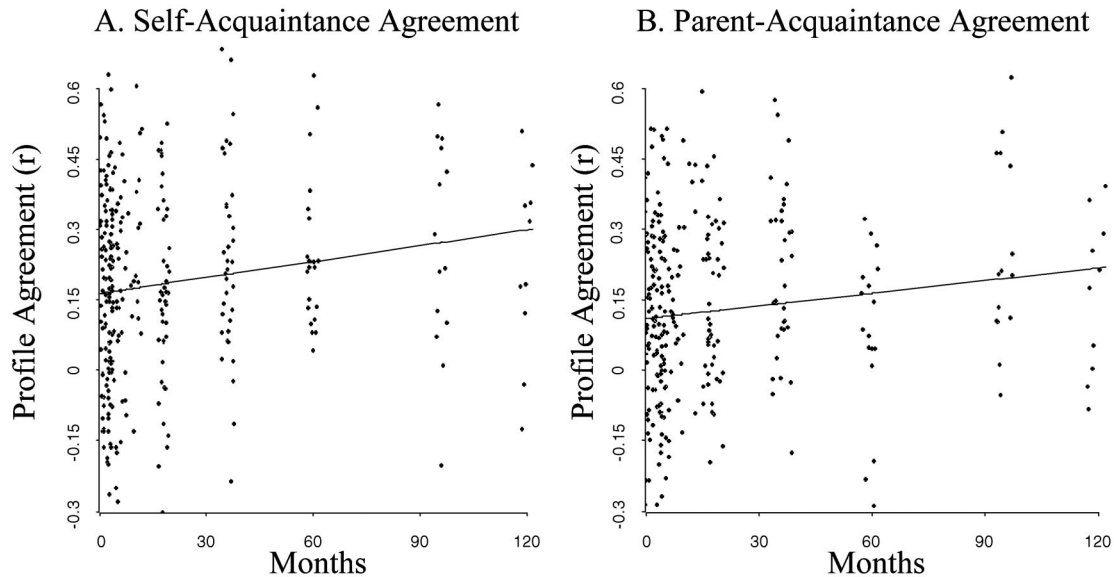
## A. Self-Acquaintance Agreement          B. Parent-Acquaintance Agreement



*Figure 3.* Relationship between length of acquaintance in months and self–acquaintance and parent–acquaintance differential accuracy (*r*) in Study 1. The data have been slightly jittered horizontally to minimize overplotting (Cohen et al., 2003).

varied across participants such that the majority of participants (58%) had ratings by relatively recent acquaintances of less than 1 year. The possibility exists that participants with longer term acquaintances differed from those with shorter term acquaintances in unknown ways, potentially leading to spurious associations with differential accuracy. Note that there is no evidence supporting such spurious associations as both self- and parent-reports on the Big Five were unrelated to length of acquaintance (all *r*s < .09, *ns*).

Second, and relatedly, it is apparent from Figure 3 that relatively few acquaintances had known the participant more than 5 years. The small percentage of participants with acquaintances of 5 or more years (14.6%) consequently had the potential for influencing the obtained results. However, the results did not change materially when the sample was restricted to those with acquaintances of fewer than 5 years.

### Study 2

Study 2 was designed to replicate Study 1 and follows the same general procedures with two major changes designed to enhance generalizability. First, we used a different assessment of the Big Five. Saucier and Ostendorf (1999) have provided lexical subcomponents (facets) to the Big Five that enable an examination of length of acquaintance both at the broad, Big Five level as well as at the more focused facet level. Second, participants were encouraged to recruit two relatively recent acquaintances (within the last year) and three long-term acquaintances (over 1 year) to serve as informants. Variability in the length of acquaintance for each participant (target person) allows an examination of changes in agreement as a function of acquaintance within each participant.

### Method

#### Participants

Introductory psychology students (*N* = 200) were recruited to participate in return for partial fulfillment of their class requirements. A total of 184 participants completed the basic study requirements of attending three measurement sessions (123 women and 61 men; mean age = 19.33 years, *SD* = 2.71). Participants provided consent for obtaining a parental rating, and a total of 153 participants had parental ratings.

Participants were encouraged to bring two short-term (within the past year) acquaintances into the laboratory and to either bring or provide contact information and consent for obtaining a rating via mail for 3 long-term (over 1 year) acquaintances in exchange for additional credit toward fulfillment of their course requirements. There were a total of 259 short-term acquaintance ratings (167 women and 92 men); mean length of acquaintance was 2.36 months (*SD* = 1.21). There were 293 long-term acquaintance ratings (215 women and 78 men; 253 obtained by mail), and mean length of acquaintance was 75.13 months (*SD* = 28.24). Of the participants, 128 had 3 or more acquaintance ratings. Compliance with the request for both short- and long-term acquaintances was excellent, and a check on this manipulation showed that the short- and long-term acquaintances differed in their length of acquaintance with the target person, *F*(1, 150) = 782.65, *p* < .0001. To further characterize the length of acquaintance, we computed the standard deviation of length of acquaintance within each participant's complete set of acquaintances (length of acquaintance in months: *M* = 36.44, *SD* = 21.48). As in Study 1, individuals who indicated that they were relatives or romantic partners of the target person were excluded from the analysis.

#### Materials, Design, and Procedure

Participants rated themselves on 100 unipolar trait adjectives: 24 for Agreeableness, 22 for Conscientiousness, 22 on Extraversion, 16 on Neuroticism, and 16 for Openness to Experience, which were extracted from Saucier and Ostendorf (1999; Table 2). Appendix B lists the specific trait adjectives used in the present study. All ratings were on a 9-point scale ranging from 0 (*extremely inaccurate*) to 8 (*extremely accurate*). As in Study 1, participant's self-rating instructions limited self-assessments of behavior to the previous week. Participants completed the self-report inventory three times, at no less than 1-week intervals, in a lecture hall reserved for that purpose. Self-assessments were aggregated across the three assessment occasions.

Acquaintances and parents rated the participant on the same 100 unipolar trait adjectives extracted from Saucier and Ostendorf (1999) by using the same 9-point rating scale. Acquaintances and parents received Goldberg's (1992) standard rating instructions, with the participant's name embedded within the instructions, and indicated how long they had known the participant, with the choices being less than 1 month, 1–2 months, 2–3 months, 3–4 months, 4–5 months, 5–6 months, 6–8 months, 8–10 months, 10–12 months, 1–2 years, 2–3 years, 3–4 years, 4–5 years, 5–6 years, 6–8 years, 8–10 years, or 10+ years. Length of acquaintance, measured in months, was determined by taking the midpoint of the interval range selected by the acquaintance with 120 months being the maximum length coded.

## Results

### Length of Acquaintance and Trait-Level Agreement

To examine the relationship between length of acquaintance and agreement, we divided acquaintances into short-term acquaintances (those who had known the participant for less than 1 year) and long-term acquaintances (those who had known the participant for over 1 year). We then calculated the correlation of ratings among short-term acquaintances and among long-term acquaintances for the Big Five and its subcomponents. We then aggregated all available short- and long-term acquaintance ratings separately to examine the relationship between self–acquaintance and parent–

acquaintance ratings for short- and long-term acquaintances. Table 1 summarizes the results of these analyses and the tests of the differences in the magnitude of the correlation coefficients for short- and long-term acquaintances. All significance tests of differences presented in Table 1 were conducted following the procedures outlined by Steiger (1980) for related-sample correlations.

Trait-level consensus among short-term and among long-term acquaintances was generally moderate to large according to Cohen's (1988) norms. The magnitude of these levels of consensus, however, did not differ significantly between short- and long-term acquaintances at either the broad Big Five level or at its subcomponent level. Similarly, agreement between self- and acquaintance ratings was mostly moderate and, although generally lower than consensus levels among acquaintances, did not differ significantly for short-term versus long-term acquaintances.

In contrast, levels of agreement between acquaintances and parents were related to length of acquaintance. Long-term acquaintances had significantly higher levels of agreement with parents than did short-term acquaintances on Agreeableness, Neuroticism, and Openness to Experience. Short-term acquaintances had very low levels of agreement with parents on these three personality dimensions and their subcomponents, whereas long-term acquaintances reached moderate levels of agreement.

Table 1

*Pearson Product–Moment Correlations Among Acquaintances and Between Acquaintance Reports and Self- and Parent-Reports for Short- and Long-Term Acquaintances in Study 2*

| Big Five trait and facet | A–A | | S–A | | P–A | |
|---|---|---|---|---|---|---|
| | Short | Long | Short | Long | Short | Long |
| Agreeableness | .45 | .40 | .20 | .23 | .07 | .35[***] |
| *Warmth* | .41 | .43 | .23 | .29 | .14 | .41[***] |
| *Gentleness* | .33 | .38 | .12 | .29 | .07 | .31[*] |
| *Generosity* | .33 | .36 | .17 | .14 | .06 | .30[*] |
| *Modesty* | .36 | .37 | .34 | .15 | .11 | .25[*] |
| Conscientiousness | .32 | .31 | .41 | .26 | .23 | .20 |
| *Orderliness* | .47 | .39 | .59 | .49 | .43 | .46 |
| *Decisiveness* | .20 | .35 | .29 | .16 | .09 | .13 |
| *Reliability* | .20 | .19 | .26 | .12 | .11 | .16 |
| *Industriousness* | .29 | .23 | .27 | .23 | .03 | .15 |
| Extraversion | .54 | .46 | .33 | .40 | .35 | .43 |
| *Sociability* | .51 | .28 | .18 | .31 | .39 | .42 |
| *Unrestraint* | .39 | .50 | .39 | .55 | .36 | .47 |
| *Assertiveness* | .32 | .44 | .14 | .16 | .17 | .21 |
| *Activity* | .41 | .24 | .40 | .36 | .20 | .31 |
| Neuroticism | .31 | .35 | .13 | .24 | .13 | .33[*] |
| *Irritability* | .34 | .43 | .18 | .26 | .14 | .37[*] |
| *Insecurity* | .15 | .23 | .12 | .23 | .13 | .28 |
| *Emotionality* | .15 | .26 | .11 | .27 | .08 | .27 |
| Openness to Experience | .32 | .24 | .14 | .20 | .00 | .33[***] |
| *Intellect* | .30 | .10 | .21 | .24 | .15 | .21 |
| *Imagination* | .23 | .35 | .21 | .31 | .09 | .36[**] |
| *Perceptiveness* | .23 | .19 | .00 | .10 | −.05 | .31[***] |
| *N* | 92 | 85 | 157 | 153 | 135 | 131 |

*Note.* Significance tests refer to the difference between the agreement correlations for short- and long-term acquaintances for that pair of raters (i.e., acquaintance–acquaintance [A–A], self–acquaintance [S–A], or parent–acquaintance [P–A]).
[*] *p* < .05.   [**] *p* < .01.   [***] *p* < .001.

## Analytic Strategy: Profile Measures

With 3 to 5 acquaintances of varying duration available for most participants, there were up to five self–acquaintance profile and parent–acquaintance correlations nested within participant. We consequently were able to model within-participant changes in profile measures (raw profile consensus, stereotype accuracy, and differential accuracy) as a function of length of acquaintance by adapting a multilevel modeling approach for these profile measures. Note that all profile measures were calculated as in Study 1. Acquaintances within participants served as the Level 1 units, and participants served as the Level 2 units. Specifically, to examine whether a particular profile measure is related to length of acquaintance, we ran the following analysis separately for self–acquaintance and parent–acquaintance profile measures:

$$z_{ij} = \beta_{0i} + \beta_{1i}\, \text{Month}_{ij} + e_{ij}. \qquad [\text{Level 1}]$$

Here $z_{ij}$ is the Fisher $z$-transformed profile measure for participant $i$ and acquaintance $j$. Fisher's $r$–$z$ transformation is used as the outcome as it stabilizes the variance of the correlation, thus making the assumption of homogeneity of variance in the errors more plausible. $\text{Month}_{ij}$ is the length of acquaintance for participant $i$ and acquaintance $j$. The coefficients $\beta_{0i}$ and $\beta_{1i}$ are the intercept and slope, respectively, for participant $i$. That is, $\beta_{0i}$ is the predicted profile measure for person $i$ at initial acquaintance (i.e., $\text{Month}_{ij} = 0$) and $\beta_{1i}$ is the predicted increase in the level of profile agreement with an acquaintance during each month of the study. Both $\beta_{0i}$ and $\beta_{1i}$ can vary randomly across participants, as is illustrated in the following equations:

$$\beta_{0i} = \gamma_{00} + u_{0i},$$
$$\beta_{1i} = \gamma_{10} + u_{1i}. \qquad [\text{Level 2}]$$

Here $\gamma_{00}$ and $\gamma_{10}$ are the estimated mean intercept and slope, respectively. The test of $\gamma_{10}$ thus represents the test whether, on average across participants, profile consensus and self–other agreement increases with length of acquaintance. Acquaintance-related covariates can be added to the Level 1 model, and stable target-related characteristics can be added to the Level 2 model that might potentially moderate the acquaintance-consensus relationship. In the Level 1 equation, we included as potential covariates the acquaintance's gender, whether or not they were living with the participant (roommates), and whether the acquaintance was also participating in the study.[4] In the Level 2 equation, we examined participant's gender and the participant's self-report on the Big Five as potential moderators. All models were estimated with hierarchical linear modeling (HLM version 5.05) with all available data under restricted maximum likelihood (Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong, & Congdon, 2001).[5]

## Length of Acquaintance and Raw Profile Correlations

Initial levels of raw profile self–acquaintance agreement and parent–acquaintance consensus were both moderate to high ($r = .29$, $p < .001$, and $r = .37$, $p < .001$, respectively). Moreover, at these initial levels, raw profile agreement and consensus do not change either significantly or appreciably with increased levels of acquaintance. In terms of correlations, self–acquaintance raw pro-

file agreement decreased slightly by .006 over 5 years, $t(467) = -.36$, $ns$, and parent–acquaintance raw profile consensus increased slightly by .014 over 5 years, $t(467) = .96$, $ns$.

## Length of Acquaintance and Stereotype Accuracy

Initial levels of stereotype accuracy in acquaintance reports were high ($r = .52$, $p < .001$).[6] Of interest, and as predicted by the WAM, stereotype accuracy declined significantly with increased levels of acquaintance such that over 5 years, stereotype accuracy dropped by .06, $t(177) = -3.055$, $p = .003$. After controlling for length of acquaintance, stereotype accuracy declined significantly (a) if the acquaintance was also a participant in the study and (b) if the acquaintance was living with the participant. The magnitude of these reductions in stereotype accuracy was $r = .14$, $t(542) = -4.02$, $p < .001$, and $r = .10$, $t(542) = 3.37$, $p = .001$, respectively.

## Length of Acquaintance and Differential Accuracy

*Acquaintance consensus.* On average there was differential accuracy both for short-term acquaintances, mean $r(98) = .17$, $p < .0001$, and long-term acquaintances, mean $r(98) = .20$, $p < .0001$. Of note, this difference between short-term and long-term acquaintances in their average levels of differential accuracy was statistically significant ($z = 2.09$, $p < .05$).

*Self–acquaintance agreement.* Length of acquaintance had a positive association with self–acquaintance agreement in differential accuracy, $t(471) = 2.05$, $p < .05$. On average, across participants, new acquaintance differential accuracy was estimated to be $r = .15$, $t(471) = 11.66$, $p < .001$. After 5 years, this level of differential accuracy rose to an estimated $r = .18$. We also probed the data for a possible nonlinear association between length of acquaintance and self–acquaintance agreement in differential accuracy. Examination of the nonparametric Lowess fit did not show visual evidence of nonlinearity. The addition of a quadratic component to the model did result in a marginally significant contribution to the overall model, $t(462) = -1.94$, $p < .06$, suggesting a possible slight deceleration of agreement over time. This finding is consistent with the WAM's prediction of an asymptotic relationship between length of acquaintance and differential accuracy. Figure 4 depicts the linear relationship between length of acquain-

---

[4] A number of participants signed up for the study with their friends and consequently served as each other's acquaintances. Participants in the study comprised 12.5% of the acquaintance reports, and we included this as a Level 1 acquaintance-varying covariate.

[5] Although there were substantial individual differences in initial profile agreement, the variance across individuals in the magnitude of the slope was only marginally significant for self–acquaintance differential accuracy profile agreement and could not be estimated for parent–acquaintance differential accuracy as this model would not converge to a solution. Consequently, we constrained this relationship to be equal for all participants for both self–acquaintance and parent–acquaintance analyses (i.e., for Level 2 we set $\beta_{1i} = \gamma_{10}$, which presumes that the rate of increase in profile agreement does not vary randomly across individuals).

[6] As in Study 1, we report the results for stereotype accuracy based on the average parental response. The results for stereotype accuracy based on self-reports were equivalent which is not surprising given that the average parental response correlates $r(98) = .80$, $p < .0001$, with the average self-report.
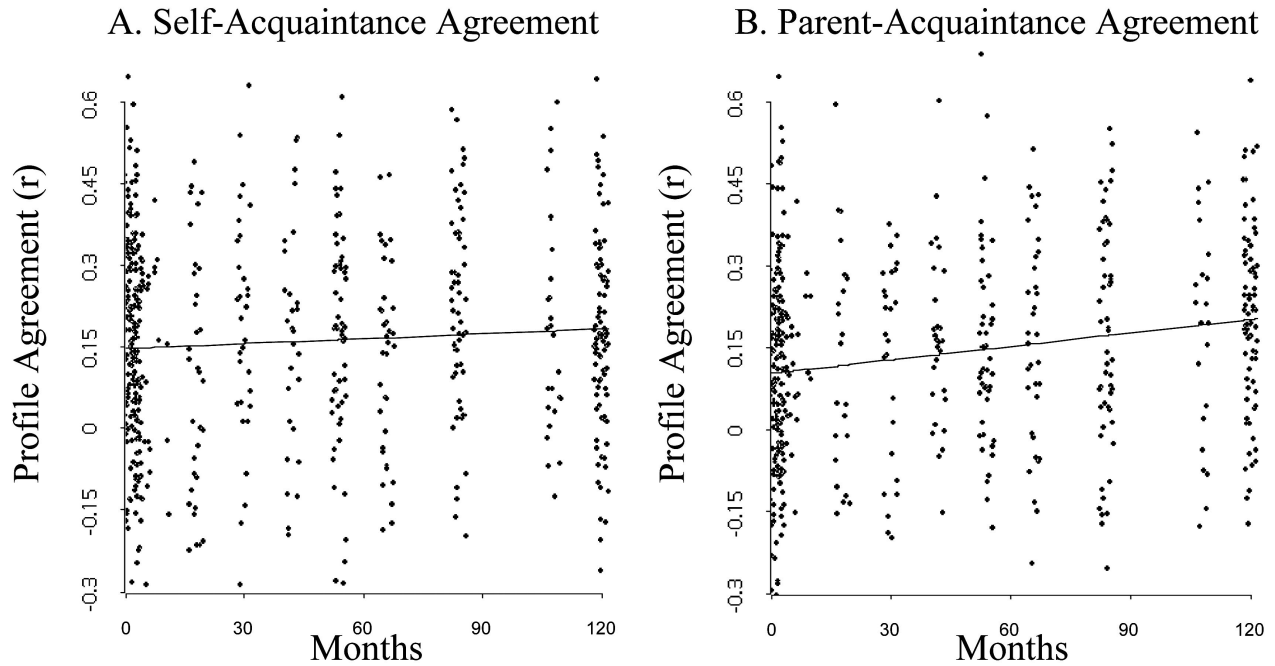
## A. Self-Acquaintance Agreement     B. Parent-Acquaintance Agreement



*Figure 4.* Relationship between length of acquaintance in months and self–acquaintance and parent–acquaintance differential accuracy (*r*) in Study 2. The data have been slightly jittered horizontally to minimize overplotting (Cohen et al., 2003).

tance and self–peer differential accuracy with all available data. In examining the set of Level 1 and Level 2 predictors, the only significant relationship that emerged was that for Conscientiousness, $t(461)=2.17$, $p < .05$. As illustrated in Figure 5, self–

acquaintance agreement in differential accuracy increased more quickly for more conscientious participants.

*Parent–acquaintance consensus.* As with self–acquaintance differential accuracy, on average, across participants, parent–
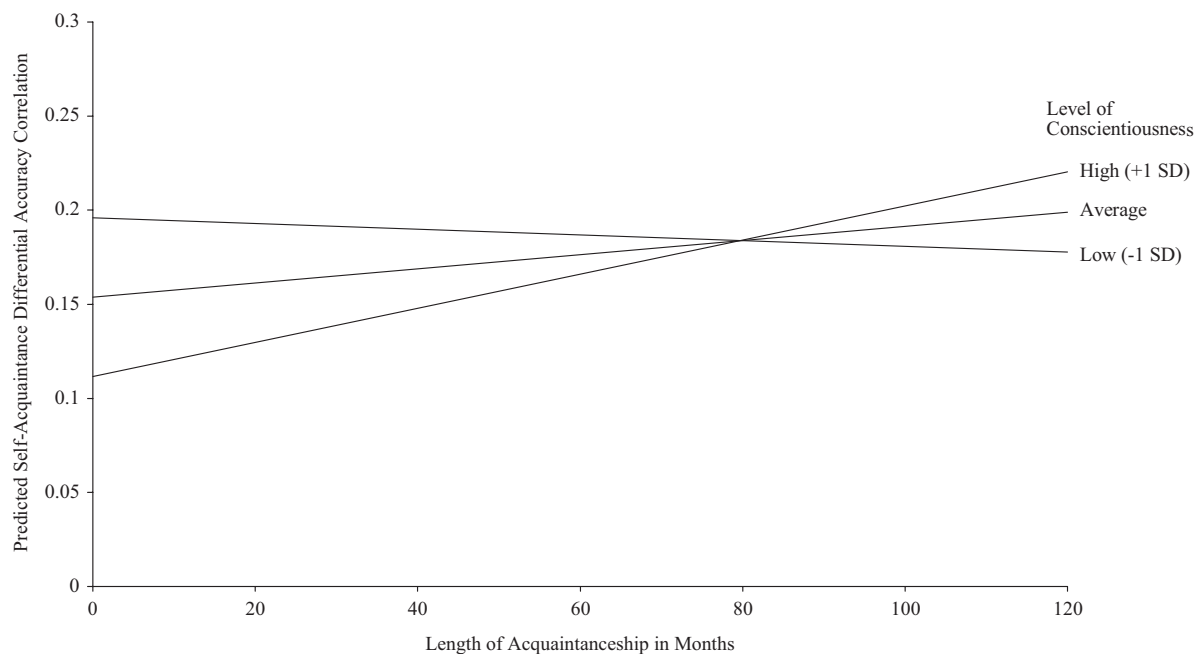


*Figure 5.* Relationship between length of acquaintance in months and self–acquaintance differential accuracy (*r*) in Study 2 moderated by self-reported level of conscientiousness.

acquaintance differential accuracy showed a linear increase with length of acquaintance, $t(471) = 3.69$, $p < .001$. The initial level of parent–acquaintance consensus in differential accuracy was estimated to be $r = .11$, $t(471) = 7.59$, $p < .001$, which rose after 5 years to an estimated $r = .16$. Even at the initial meeting (Month = 0), the model estimates that acquaintance reports correspond with those from parents and this level of consensus increases with length of acquaintance. Examination of the Lowess fit and the contribution of the quadratic component, $t(462) = -1.15$, ns, did not show any evidence of a nonlinear relationship.

Examining the same set of Level 1 and Level 2 potential predictors of parent–acquaintance consensus in differential accuracy revealed the same pattern of results. Parent–acquaintance consensus increased more quickly for the more conscientious participants, $t(461)=2.77$, $p < .01$ (see Figure 6). No other variable was significantly related to either profile agreement or the rate of change in profile agreement (all $ts < 1.7$, ns).

## Discussion

For trait-level ratings, length of acquaintance was not significantly related to consensus among acquaintances or to self–acquaintance agreement. In contrast, length of acquaintance was related to parent–acquaintance consensus for three of the Big Five dimensions. It is interesting to note that the two traits that did not show an effect, Extraversion and Conscientiousness, are those dimensions on which observers demonstrate consensus with only minimal information (e.g., Albright et al., 1988; Borkenau & Liebler, 1992). Achieving consensus on Agreeableness, Neuroticism, and Openness to Experience requires a (much) longer observational window as the relevant behavioral cues used to form accurate impressions may be more sparsely distributed over time.

This longer required observational window may consequently provide the opportunity to model the development of agreement and thereby detect the relationship between length of acquaintance and agreement.

Length of acquaintance was again not significantly related to raw profile consensus or self–other agreement. However, stereotype accuracy declined significantly with increased acquaintance, consistent with the WAM's predictions. It is interesting to note that both participation in the study (i.e., dyads that signed up for Study 2 jointly and served as one of each other's acquaintances) and living with the participant resulted in lower levels of stereotype accuracy. Speculatively, these variables may serve as indicators of having much more extensive knowledge of the participant. Living with someone certainly provides a wider and richer observational window than is available to casual acquaintances. Having witnessed more acts, according to WAM, would lead to a reduction in level of stereotype accuracy, and time (length of acquaintance), cohabitation, and dyads that decide to jointly enroll in a study might all serve as indirect measures of the number of witnessed acts.

Replicating the results of Study 1 with a different assessment instrument, both self–acquaintance and parent–acquaintance differential accuracy increased with length of acquaintance. Although recent acquaintances agreed with self-reports and with parent-reports, long-term acquaintances were better able to agree with both self-ratings and parent ratings on the relative order of attributes within participants than were more recent acquaintances.

Of note, both self–acquaintance and parent–acquaintance differential accuracy increased more rapidly as a function of the length of acquaintance for more conscientious participants. According to the WAM, this relationship would occur if conscientiousness were
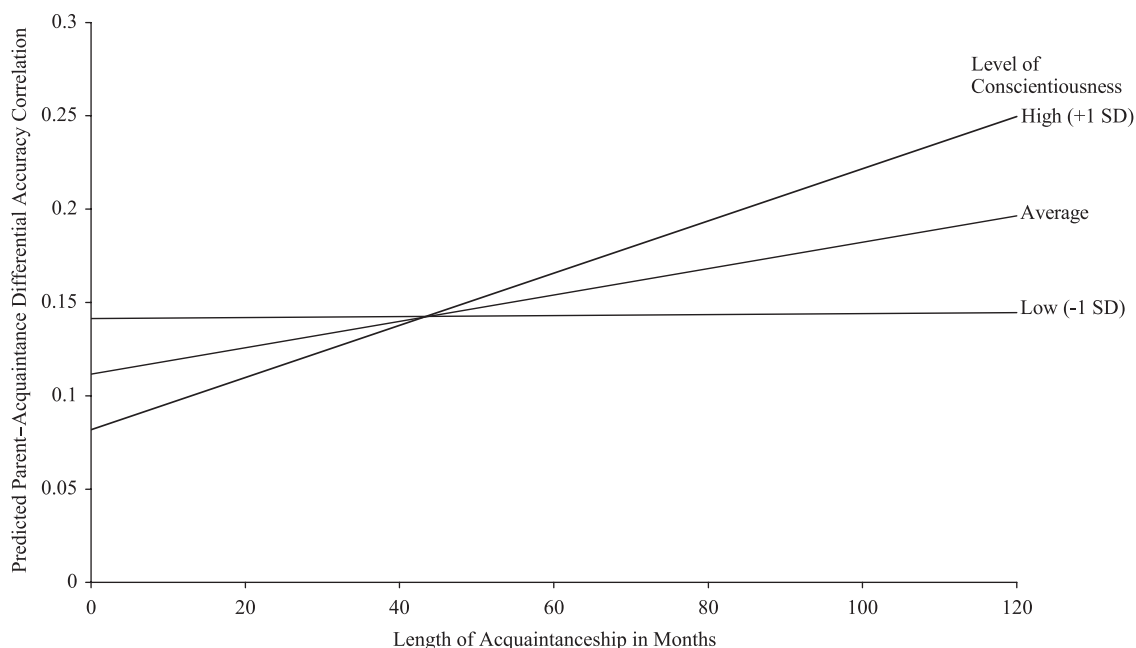


*Figure 6.* Relationship between length of acquaintance in months and parent–acquaintance differential accuracy ($r$) in Study 2 moderated by self-reported level of conscientiousness.

related to behavioral consistency and/or more diagnostic behavioral cues (i.e., greater levels of shared meaning, $\rho_2$). Under these conditions, overall levels of eventual (asymptotic) agreement would be greatest for individuals high in conscientiousness, resulting in a stronger relationship between length of relationship and consensus and self–other agreement in differential accuracy than for less conscientious individuals. Further research exploring the precise behavioral mechanisms underlying this relationship is needed.

## General Discussion

Do we come to know an individual's personality better with further contact? Drawing on insights from Funder's (1995, 1999) RAM model, we used Kenny's (1991, 1994) WAM model to identify conditions under which increased length of acquaintance may be related to increased consensus and self–other agreement. We also extended the WAM model to predict raw profile consensus and self–other agreement in conjunction with Cronbach's (1955) components of stereotype accuracy and differential accuracy. As predicted by the extended WAM model, increased length of acquaintance led to greater differential accuracy, no change in raw profile correlations, and reduced stereotype accuracy. In other words, impressions cohere and develop over time such that correspondence on relative patterns of behavior does emerge across time and relies less on how people behave in general. In contrast, and consistent with previous theory and research, across two studies we did not find strong evidence of enhanced consensus or self–other agreement across self-, acquaintance-, and parental-reports of mean trait levels as a function of length of acquaintance.

### The Elusive Length of Acquaintance Effect

Our naive intuition about the effects of length of acquaintance on judgments of personality is supported by cross-sectional research such as, for example, that of Watson et al. (2000) who demonstrated that married couples have higher levels of self–other agreement than do dating couples or friendship dyads. Yet such cross-sectional research cannot rule out potential selection biases such as, for instance, systematic differences between married couples and other dyads that, in addition to length of acquaintance, are related to the level of agreement. Indeed, substantial longitudinal empirical research—which can help rule out many potential alternative explanations of the length of acquaintance effect—has generally not produced evidence confirming this elusive length of acquaintance effect. Given (a) such a stark juxtaposition between research findings with these different study designs and (b) Kenny's (1991, 1994) WAM model that predicts that consensus quickly asymptotes under many conditions, it would seem reasonable to conclude that the results of cross-sectional studies must be driven by as-of-yet undetermined confounding factors.

However, it may be premature to conclude that consensus and self–other agreement does not change measurably with enhanced contact. Previous cross-sectional and longitudinal studies have generally differed in two fundamentally important ways. First, the longitudinal research, with limited exceptions (e.g., Albright, 1990; L. Albright, personal communication, January 12, 2004; Park et al., 1997), has been conducted in controlled circumstances

such as in classrooms, experimental laboratory groups, or other groups that meet infrequently for contact and assessment purposes. As a function of exerting experimental control over the acquaintance process, the range of different situations encountered and, consequently, the variability of behaviors that the reporters observe is by design highly constrained. As the range of different observed situations is restricted, consistency within an observer ($\rho_1$) will by necessity increase. These are precisely the conditions under which length of acquaintance will have no appreciable impact on levels of consensus according to the WAM model. In contrast, naturally occurring dyads (e.g., acquaintances, friends, couples) provide the opportunity to observe the target person in different and more diverse environments. As a result, the observed behaviors will be more variable, resulting in lower average correlations among scale weights across different observed behaviors (i.e., low values of $\rho_1$). Thus we would expect to see an effect of length of acquaintance on consensus and self–other agreement emerge among naturally occurring acquaintances that do see behavior in many different situations.

Second, longitudinal research on acquaintance to date has been conducted by using relatively short-term studies—all have been under 1 year in length with the modal design being 1 semester or less in duration. It is quite plausible that this length of time is insufficient to reliably detect changes in agreement as a function of length of acquaintance. The present studies, in contrast, demonstrate that differential accuracy increases with length of acquaintance over a period of 10 years. Indeed, Watson et al.'s (2000) study showed that couples married for 17 years on average had trait-level correlations only .15 higher than friendship dyads who had known each other fewer than 3 years on average. This result suggests an estimated effect size increase of $r = .05$ for every 5 years, a value which is consistent with that observed in the present studies. If the observational window were restricted to the 1st year of the acquaintance, it is clear that the statistical power to detect the observed length of the acquaintance-consensus relationship would have been drastically reduced. This suggests that our naive intuition that agreement in judgments of personality increases with time may indeed be correct but that the process is relatively slow in maturing, and the magnitude of the effect, though modest on an annual basis, does accumulate over long periods of time.

### Differential Accuracy, Stereotype Accuracy, and the Length of Acquaintance Effect

The use of differential accuracy measures under these conditions—long observational windows and naturally occurring dyads—conveys theoretical advantages over trait-level analyses to detect the length of acquaintance effect. Trait-level analyses can potentially demonstrate the length of acquaintance effect as in Study 2 when peer–parent consensus increased as a function of acquaintance for Agreeableness, Neuroticism, and Openness to Experience. At a practical level, however, differential accuracy measures may benefit from two sources. First, differential accuracy measures aggregate all available data across traits in a single analysis, leading to increased statistical power relative to trait-level analyses. Second, differential accuracy measures remove elevation and differential elevation components that are presumably unrelated to acquaintance and would only serve to attenuate that relationship.

The lack of a relationship between raw profile correlational measures of self–other agreement and consensus and length of acquaintance, as Study 2 illustrates, masks two different trends—a decrease in stereotype accuracy, coupled with an increase in differential accuracy—that essentially cancel each other out. This demonstrates that even as raw levels of (profile) consensus remains unchanged, impressions become more individuated and accurate with increased levels of acquaintance.

Finally, the interpretation of the present results can be viewed from several different perspectives. In Study 2, we demonstrated that differential accuracy for recently acquainted peers was significantly lower than for long-term peers. In comparison, the magnitude of the relationship between length of acquaintance and self–acquaintance and parent–acquaintance differential accuracy was substantially stronger. Self–other agreement can be viewed as an index of accuracy (e.g., see Funder & Colvin, 1997; Funder & West, 1993; Paulhus & Reynolds, 1995) and parent–acquaintance consensus can be viewed in a similar light. According to the WAM (see Kenny, 1991), accuracy will increase with length of acquaintance.[7] Viewed in this manner, the present results are congruent with the predictions derived from WAM. Nonetheless, WAM also predicts that consensus will also increase with length of acquaintance under the right conditions and these conditions are better approximated under naturally occurring dyadic relationships than in dyads that comprise groups in the laboratory.

## Examining Length of Acquaintance: Methodologies and Limitations

The present studies demonstrate that self–other differential accuracy agreement and consensus between informants in impressions of personality profiles increases with length of acquaintance, yet the present results rest on a study design that is cross-sectional. However, in Study 2 we introduced a design improvement by deliberately selecting multiple peers of varying length of acquaintance. This design feature eliminates the major confound present in typical cross-sectional study—namely, that participants with longer term acquaintances may differ systematically from participants with shorter term acquaintances in ways that are related to agreement. Yet, the possibility remains that longer term acquaintances qualitatively differ from shorter term acquaintances. For example, it is possible that some of the longer term acquaintances may have some overlap with the target person's parents in their observations of the participant. Although this might result in enhanced parent–acquaintance consensus for longer term acquaintances such as that observed in the trait-level agreement in Study 2, it would not explain the relationship between length of acquaintance and self–peer or peer–peer differential accuracy. Alternatively, the methodology of deliberately sampling shorter term versus longer term acquaintances may have resulted in different types of acquaintances in these two general groups. We note that the monotonic and apparently continuous relationship between length of acquaintance and differential accuracy across the range of acquaintance—that is, length of acquaintance is related to agreement within the longer-term acquaintances—argues against such qualitative group differences.

At first glance, it would seem that the ideal study design to examine the length of acquaintance effect is the classic longitudinal design. However, the study of naturally occurring dyads from the point of initial acquaintance over several years (e.g., a decade) presents its own interpretational and logistic problems. Over the course of several years, the acquaintances who have maintained close contact might differ systematically from those who chose not to continue the relationship or who became geographically separated, presenting interpretational difficulties. Logistically, most participants would be unable to correctly forecast which of a set of new acquaintances at the initiation of the study would continue as acquaintances for the full duration of a multiyear study. A large proportion of the participants could be expected to have no informants at the completion of the study. The selection of research contexts that may minimize attrition (e.g., the workplace) may once again constrain the diversity of situations in which the target's behavior is observed. Thus, in practice, the theoretically ideal longitudinal study that could optimally study the development of measures of consensus and self–acquaintance agreement would be very difficult to implement in practice, except perhaps in the unusual context of an isolated small community with a stable population.

Despite these interpretational and logistic difficulties, a classic longitudinal study could potentially provide important data to estimate the form of the relationship between length of acquaintance and agreement. Such studies are easiest to implement in the initial stages of acquaintance in which the role of stereotypes and expectations (i.e., the unabridged WAM and PERSON; see Kenny, 1994, 2004) could impact consensus and self–other agreement— which could account for the higher levels of consensus among acquaintances in Table 1—and where we would expect strong nonlinearity in the relationship between agreement and length of acquaintance (e.g., Borkenau et al., 2004).

## Summary

Differential accuracy increased and stereotype accuracy decreased with length of acquaintance—consistent with predictions derived from Kenny's (1991, 1994) WAM—whereas raw profile correlations and trait-level relationships did not change appreciably. The ability to detect these relationships emerged, we argue, because of (a) theoretical differences in differential accuracy versus trait-level and raw profile consensus and stereotype accuracy, (b) the use of naturally occurring dyads, and (c) the examination of the length of acquaintance effect over a long span. Over long periods of time of observing someone in many different situations, as we typically do with our friends and acquaintances, our impressions of their overall personality—the relative ordering of their different personality attributes—coheres in a manner congruent with other individuals' impressions and relies less on how people are in general. In short, there is evidence that we do come to know others better over time.

---

[7] The index of accuracy defined in Kenny (1991), based on the generalizability theory framework, is obtained by taking the square root of the agreement correlation. Under this framework, the correlation itself, $\hat{\rho}$, is interpreted as the percentage of shared variance (see Ozer, 1985, for an explanation of these different frameworks) that necessitates taking the square root in order to obtain the validity coefficient. Taking the square root of Equation 1, for example, essentially stretches out the scale of time (length of acquaintance) resulting in a relationship between accuracy and length of acquaintance (e.g., see Kenny, 1991, Figure 4).

## References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Albright, L. (1990). [A longitudinal study of consensus and accuracy in interpersonal perception]. Unpublished raw data.

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., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69,* 518–529.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111,* 256–274.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self–other agreement. *Personality and Social Psychology Bulletin, 20,* 367–378.

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. (2004). Toward understanding assessments of the Big Five: Multitrait–multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality, 72,* 845–876.

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.

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. (1992). Implicit personality theory and the five-factor model. *Journal of Personality, 60,* 295–328.

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

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner. A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86,* 599–614.

Chaplin, W. F., & Panter, A. T. (1993). Shared meaning and the convergence among observers' personality descriptions. *Journal of Personality, 61,* 553–585.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatter plots. *Journal of American Statistical Association, 74,* 829–836.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. A. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Colvin, C. R., & Funder, D. C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology, 60,* 884–894.

Cook, R. D., & Weisberg, S. (1999). *Applied regression including computing and graphics.* New York: Wiley.

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

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles.* New York: Wiley.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51,* 360–392.

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 Press.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55,* 149–158.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology, 60,* 773–794.

Funder, D. C., & Colvin, C. R. (1997). Congruence of self and others' judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). Orlando, FL: Academic Press.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52,* 409–418.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69,* 656–672.

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

Goldberg, L. R. (1992). The development of markers for the Big Five factor structure. *Psychological Assessment, 4,* 26–42.

Jackson, D. N., Neill, J. A., & Bevan, A. R. (1973). An evaluation of forced-choice and true-false item formats in personality assessment. *Journal of Research in Personality, 7,* 21–30.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98,* 155–163.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford Press

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Psychological and Social Psychology Review, 8,* 265–280.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102,* 390–402.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin, 116,* 245–258.

Kenny, D. A., & Winquist, L. A. (2001). The measurement of interpersonal sensitivity: Consideration of design, components, and unit of analysis. In J. Hall & F. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement (*pp. 265–302). Englewood Cliffs, NJ: Erlbaum.

Kurtz, J. E., & Sherker, J. L. (2003). Relationship quality, trait similarity, and self–other agreement on personality ratings in college roommates. *Journal of Personality, 71,* 21–48.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology, 65,* 1178–1187.

Malloy, T. E., Agatstein, F. Yarlas, A., & Albright, L. (1997). Effects of communication, information overlap, and behavioral consistency on consensus in social perception. *Journal of Personality and Social Psychology, 73,* 270–280.

McCrae, R. R., Terracciano, A., & Personality Profiles of Cultures Project. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89,* 407–425.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4,* 681–691.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307–315.

Park, B., DeKay, M. L., & Kraus, S. (1994). Aggregating social behavior into person models: Perceiver-induced consistency. *Journal of Personality and Social Psychology, 66,* 437–459.

Park, B., & Judd, C. M. (1989). Agreement on initial impressions: Differences due to perceivers, trait dimensions, and target behaviors. *Journal of Personality and Social Psychology, 56,* 493–505.

Park, B., Kraus, S., & Ryan, C. S. (1997). Longitudinal changes in

consensus as a function of acquaintance and agreement in liking. *Journal of Personality and Social Psychology, 72,* 604–616.

Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology, 63,* 816–824.

Paulhus, D. L., & Reynolds, S. (1995). Enhancing target variance in personality impressions: Highlighting the person in person perception. *Journal of Personality and Social Psychology, 69,* 1233–1242.

Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of Personality and Social Psychology, 56,* 823–833.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. Jr. (2001). *HLM 5: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology, 76,* 613–627.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Story, A. L. (2003). Similarity of trait construal and consensus in interpersonal perception. *Journal of Experimental Social Psychology, 39,* 364–370.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology, 57,* 120–128.

Watson, D., & Clark, L. A. (1991). Self- versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology, 60,* 927–940.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78,* 546–558.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review, 1,* 203–222.

# Appendix A

## Derivation of the Weighted Average Model for Differential Accuracy

Below we extend Kenny's (1991) weighted average model to differential accuracy. Let $\mathbf{x}$ be the vector of an observer's impressions of a target individual across $p$ traits (after subtracting the stereotype effect of the average person's profile), defined as follows:

$$\mathbf{x} = \mathbf{S1}' + \mathbf{k}. \qquad (A1)$$

Here $\mathbf{S}$ is the $p$ by $n$ matrix of scale weights, $\mathbf{k}$ contains the $p$ by 1 vector of unique impressions for each trait, and $\mathbf{1}$ is a 1 by $n$ vector of 1s. We make three assumptions to simplify the derivation. First, the weight of the unique impression ($k$) is the same for each trait. Second, the variance of the scale weights across acts is the same for each trait and is equal to 1. Third, the vector $\mathbf{x}$ is mean centered across traits. These last two assumptions provide the units and origin of the metric for the scale weights. The choice of units and origin is arbitrary and has no impact on the resulting model or predictions.

Consider two observers A and B who observe $qn$ acts of the same target in common, where $n$ is the number of observed acts and $q$ is the proportion observed in common. The differential accuracy correlation of their impressions across $p$ traits of the target individual is as follows:

$$r_{AB} = \frac{\mathbf{x}'_A \mathbf{x}_B}{\sqrt{\mathbf{x}'_A \mathbf{x}_A} \sqrt{\mathbf{x}'_B \mathbf{x}_B}} . \qquad (A2)$$

Substituting in Equation A1 for both observers results in

$$r_{AB} = \frac{(\mathbf{1S}'_A + \mathbf{k}'_A)(\mathbf{S}_B \mathbf{1}' + \mathbf{k}_B)}{\sqrt{\mathbf{1S}'_A \mathbf{S}_A \mathbf{1}' + \mathbf{k}'_A \mathbf{k}_A} \sqrt{\mathbf{1S}'_B \mathbf{S}_B \mathbf{1}' + \mathbf{k}'_B \mathbf{k}_B}}$$

$$= \frac{\mathbf{1}'_A \mathbf{S}'_A \mathbf{S}_B \mathbf{1}'}{\sqrt{\mathbf{1S}'_A \mathbf{S}_A \mathbf{1}' + pk^2} \sqrt{\mathbf{1S}'_B \mathbf{S}_B \mathbf{1}' + pk^2}} . \qquad (A3)$$

Note that the uniquenesses for each observer are assumed to be independent (i.e., $\mathbf{k}'_A \mathbf{k}_B = 0$), uncorrelated across traits, and that

the squared uniquenesses for each observer are the same (i.e., $k_A^2 = k_B^2 = k^2$). To evaluate this profile correlation further, we consider the constituent elements of the $n$ by $n$ symmetric matrices $\mathbf{S}'_A \mathbf{S}_A$, $\mathbf{S}'_B \mathbf{S}_B$, and $\mathbf{S}'_A \mathbf{S}_B$ asymptotically. Both $\mathbf{S}'_A \mathbf{S}_A$ and $\mathbf{S}'_B \mathbf{S}_B$, after dividing by the constant $p$, are matrices containing 1s on the diagonal (given the second assumption) and $\rho_1$ for off-diagonals elements as seen in Equation A4. These matrices each contain a total of $n$ 1s and $(n^2 - n)$ of the $\rho_1$ correlations:

$$\frac{1}{p}\mathbf{S}'_A \mathbf{S}_A = \frac{1}{p}\mathbf{S}'_B \mathbf{S}_B = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \rho_1 & \cdots & \rho_1 & 1 \end{bmatrix}. \qquad (A4)$$

Equation A4 makes transparent that for profile consensus, $\rho_1$ represents the profile correlation across the $p$ traits of the scale weights from different acts observed by the same judge. We now examine the matrix $\mathbf{S}'_A \mathbf{S}_B$, ordering it such that the $qn$ acts observed in common are contained in the upper left quadrant.

$$\frac{1}{p}\mathbf{S}'_A \mathbf{S}_B = \begin{bmatrix} \begin{bmatrix} \rho_2 & \rho_3 & \rho_3 \\ \rho_3 & \ddots & \rho_3 \\ \rho_3 & \rho_3 & \rho_2 \end{bmatrix} & \begin{bmatrix} \rho_3 & \cdots & \rho_3 \\ \vdots & \ddots & \vdots \\ \rho_3 & \cdots & \rho_3 \end{bmatrix} \\ \begin{bmatrix} \rho_3 & \cdots & \rho_3 \\ \vdots & \ddots & \vdots \\ \rho_3 & \cdots & \rho_3 \end{bmatrix} & \begin{bmatrix} \rho_3 & \cdots & \rho_3 \\ \vdots & \ddots & \vdots \\ \rho_3 & \cdots & \rho_3 \end{bmatrix} \end{bmatrix}. \qquad (A5)$$

In Equation A5, $\rho_2$ indexes shared meaning and represent the scale weight profile correlation across the $p$ traits between two observers of the same act. These are contained on the main diagonal of the upper left quadrant of the $qn$ observed acts in common. The correlation $\rho_3$ represents the scale weight profile correlation between observers who witness different acts. The

matrix $\mathbf{S}'_A\mathbf{S}_B$ contains a total of $qn$ of the $\rho_2$ correlations and $(n^2 - qn)$ of the $\rho_3$ correlations.

Substituting in Equations A4 and A5 into Equation A3 and dividing by the constant $p$ results in Equation A3 as follows:

$$\rho_{AB} = \frac{qn\rho_2 + (n^2 - qn)\rho_3}{(n^2 - n)\rho_1 + n + k^2}.$$

## Comparing $\rho_1$ Between Trait-Level and Differential Accuracy Analyses

To illustrate how $\rho_1$ changes from the variable (e.g., trait) to a differential accuracy profile analysis, let $\sigma_s^2$ be the variance in an observer's scale weights across acts for a single trait and let $\rho_{1_{trait}}$ be the association among these scale weights for a single trait across acts. In other words, $\rho_{1_{trait}}$ is defined exactly as originally formulated by Kenny (1991) as the consistency of an observer's scale weights across acts. To simplify matters considerably, assume that both $\rho_{1_{trait}}$ and $\sigma_s^2$ remain constant across traits. Under this assumption, the covariance between scale weights across acts, both for profile as well as trait-level analyses, is $\rho_{1_{trait}}\sigma_s^2$.

Let $\rho_s$ be the correlation between scale weights for the same act between two different traits within the same observer. An association among scale weights across traits may arise from implicit personality theory as when, for example, an observer gives posi-

tive scale weights to the traits agreeableness and extraversion when simply seeing a target individual smiling. To the extent that scale weights are positively related within the same act across traits, the expected variance across the scale weights for the same act is inflated by $(p - 1)\rho_s\sigma_s^2$. This inflated variance, relative to the variance of scale weights across acts for a single trait, results in a lower $\rho_1$ correlation for differential accuracy as compared to trait-level analyses as seen in Equation A6 below:

$$\rho_{1_{DAr}} = \frac{\rho_{1_{trait}}}{1 + (p - 1)\rho_s}. \tag{A6}$$

Thus to the extent that scale weights assigned to different traits within an observer are related to each other within a single act, consistency within an observer $\rho_1$ will differ for differential accuracy as compared to trait-level analyses. This lower $\rho_1$ correlation for differential accuracy analyses implies that consensus is predicted to emerge more slowly than for trait-level analyses. Note that we assume here that the association between scale weights across traits within an observer is greater that the association between scale weights across traits across different observers (i.e., there is method variance; see Biesanz & West, 2004) and consequently the impact will be greater for $\rho_1$ than for $\rho_3$. This will consequently result in an increased ability to detect the relationship between acquaintanceship and consensus and self–other agreement for analyses examining differential accuracy.

## Appendix B

### Specific Trait Adjectives Extracted From Saucier and Ostendorf (1999; Table 2) Used in Study 2

| Trait subcomponent | Trait adjective |
| --- | --- |
| Agreeableness | |
| *Warmth* | warm, affectionate, sentimental, sensitive, unsympathetic[a], insensitive[a] |
| *Gentleness* | agreeable, cordial, antagonistic[a], hard[a], rough[a], combative[a] |
| *Generosity* | generous, charitable, helpful, greedy[a], selfish[a], stingy[a] |
| *Modesty* | modest, boastful[a], conceited[a], snobbish[a], vain[a], egocentric[a] |
| Conscientiousness | |
| *Orderliness* | organized, orderly, neat, disorderly[a], sloppy[a] |
| *Decisiveness* | decisive, firm, consistent, inconsistent[a], scatterbrained[a], illogical[a] |
| *Reliability* | reliable, dependable, responsible, prompt, unreliable[a], punctual |
| *Industriousness* | industrious, ambitious, purposeful, negligent[a], lazy[a] |
| Extraversion | |
| *Sociability* | sociable, withdrawn[a], cheerful, seclusive[a], merry |
| *Unrestraint* | talkative, untalkative[a], verbal, shy[a], reserved[a], aggressive |
| *Assertiveness* | assertive, direct, cowardly[a], straightforward, submissive[a], helpless[a] |
| *Activity* | active, daring, competitive, adventurous, uncompetitive* |
| Neuroticism | |
| *Irritability* | undemanding[a], uncritical[a], temperamental, impatient, defensive |
| *Insecurity* | relaxed[a], unstable, nervous, envious, jealous |
| *Emotionality* | emotional, unemotional[a], excitable, anxious, fidgety, suggestible |
| Openness to Experience | |
| *Intellect* | intelligent, intellectual, philosophical analytical, knowledgeable, unintellectual* |
| *Imagination* | imaginative, creative, inventive, artistic, clever, unimaginative[a] |
| *Perceptiveness* | perceptive, insightful, unobservant[a], shortsighted[a] |

[a] Reverse coded.