Running Head: INFORMATION AND PERSONALITY JUDGMENT

Quantity and Quality of Available Information Affect

the Realistic Accuracy of Personality Judgment

Tera D. Letzring

University of California, Riverside

Shannon M. Wells

Sanford Systems

David C. Funder

University of California, Riverside

Triads of previously unacquainted college students rendered judgments of each others' personalities after interacting in one of five experimental conditions that varied across three levels of information quantity and three levels of information quality. Separately, personality ratings of the members of the triad were obtained from knowledgeable informants (the self, everyday acquaintances, and clinician-interviewers), and the average of these ratings was used as a criterion for realistic accuracy. Results showed that realistic accuracy increased along with both the quantity and quality of the information available in the acquaintanceship context, and objective knowledge about the targets of judgment increased likewise. Inter-judge consensus and self-other agreement followed a similar pattern. These findings are consistent with expectations based on the Realistic Accuracy Model (RAM, Funder, 1995, 1999), and findings concerning consensus were also consistent with expectations of the Weighted Average Model (Kenny, 1994).

Quantity and Quality of Available Information

Affect the Realistic Accuracy of Personality Judgment

Every day, people make personality judgments that are based on different amounts and types of acquaintanceship, and these judgments vary in how accurate they are. Judgments may be made on the basis of brief or long-term acquaintance or, holding amount of acquaintance constant, on the basis of observations or interactions in different – and perhaps differentially revealing – situational contexts. Several factors have been hypothesized to be related to the accuracy with which personality judgments are made, including qualities of the judge, the target, the trait, and the type of information on which the judgment is based (Allport, 1937; Colvin, 1993; Funder, 1995; Vernon, 1933). The current paper reports the findings of an extensive research project designed to focus on this final factor, which has been referred to as *good information* (Funder, 1995).

Good information has two aspects: *information quantity*, the sheer amount of information that is available, and *information quality*, the degree to which the available information is relevant to personality. Information quantity has received a fair amount of research attention, and the main conclusion has been that, all other things being equal, to know someone longer is to know him or her better. Information quality has received less attention, although the available evidence suggests that not all information is created equal and that some kinds of information are more likely to be related to judgment accuracy than others (Anderson, 1984). The Realistic Accuracy Model (RAM; Funder, 1995; 1999) predicts that the availability of both more information and higher quality information will increase the probability that a personality judgment will achieve a high degree of accuracy.

The goal of the current investigation is to examine how information quantity, also referred to as *acquaintance* (Colvin & Funder, 1991; Kenny, 1994), and information quality affect three indicators of judgment accuracy: *realistic accuracy, consensus,* and *self-other agreement*. Realistic accuracy refers to a hypothetical construct representing the level of agreement between a personality judgment and what a target is really like. This construct is not directly measured by any single personality or behavior rating, but is approached to the degree that multiple means of measurement are used and combined. In the current study, we attempt to build an accuracy criterion that is as close as possible to what a person is really like by using a composite rating based on judgments of close acquaintances, clinical judges, and the self. Consensus and self-other agreement can be measured more directly than realistic accuracy, and are more commonly studied. Consensus is the level of agreement between the personality judgments rendered by two or more people about another person, which is not necessarily related to what the person is actually like. Self-other agreement is the level of agreement between judgments made about another person and that person's self-judgments, which often have been implicitly assumed to reflect reality. Indeed, many investigations refer to self-other agreement simply as *accuracy*. We report findings concerning all three indicators of judgment for two reasons. First, realistic accuracy, while relevant to consensus and self-other agreement, is far from synonymous with either of these operationalizations and potentially could yield different results. Second, consensus and self-other agreement have been employed in many prior studies (e.g., Anderson, 1984; Blackman & Funder, 1998; Kenny, Albright, Malloy, & Kashy, 1994), and our presentation of them will allow for comparison with the existing literature.

Information quantity will be varied across three experimental conditions that are similarly structured and differ only in the sheer amount of acquaintanceship they afford (minimal

interaction, 50 minutes, and 3 hours). The goal is to add to the growing literature on the acquaintanceship effect by replicating its existence in a new context and across different levels of information quantity, and by examining the effect upon realistic accuracy as well as consensus and self-other agreement.

The more distinctive contribution of the present research is to offer one of the first empirical studies to directly examine the relationships between experimentally-manipulated *information quality* and realistic accuracy, consensus, and self-other agreement (for an example employing the criterion of self-other agreement see Anderson, 1984). The concept of information quality is complex, but at its core refers simply to the likelihood that different contexts of acquaintanceship, of equal time length, might vary in the degree to which they afford accurate judgment of individual differences in personality.   Most obviously, a context in which everybody acts the same would yield only low-quality information for personality judgment, because it would not afford information about individual differences in personality, whereas one where behavior is freer to vary would yield higher-quality information. It is also possible that different kinds of information – e.g. thoughts and feelings as opposed to hobbies and activities (Anderson, 1984) – are differentially informative about personality. It is difficult to examine information quality independent of information quantity in natural settings of acquaintanceship because people may feel more behaviorally free and share more personality-relevant information with acquaintances they have known longer. The current study is designed to experimentally and independently manipulate the quantity and quality of information available to judges of personality as they meet and interact with their targets in a laboratory context.

Criteria for Accuracy

In order to examine the accuracy of a personality judgment, it is first necessary to select or construct a criterion to represent what the person in question is really like, which is no simple matter. The difficulty of the criterion problem has led many researchers to shy away from studying the accuracy of judgments of real people and to turn their attention instead to models of the cognitive processing of artificial stimuli (Funder, 1987, 1995; Jones, 1985). One possible way to deal with the criterion problem – besides changing research topics – is to gather diverse information about the target of judgment from multiple sources, and to combine this information into a broadly-based description that may begin to approach what the person is really like (Funder, 1995). A less-expensive alternative is to study outcomes that are assumed to approximate realistic accuracy, such as consensus and self-other agreement (Ambady, Hallahan, & Rosenthal, 1995; Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Funder, 1980; Funder, Kolar, & Blackman, 1995; John & Robins, 1993; Kenny, 1991).

If a researcher is concerned solely with consensus, the need for a further accuracy criterion can by bypassed because the level of agreement among judges, regardless of the accuracy of the judgments, is itself the phenomenon of interest. Still, it is important to remember that two or even many judges can agree with each other (and have high consensus) and at the same time be wrong about the personality of the target (and have low accuracy; Funder & Colvin, 1988; Funder & West, 1993; Kenny, 1991). The study of self-other agreement similarly bypasses the issue of realistic accuracy by using the target's rating of himself or herself as the criterion, and assuming that the judgment of another is accurate to the degree that it converges with the self-rating. Still, self-other agreement does not guarantee highly accurate judgments

because the self-rating will not adequately represent personality if people are unwilling or unable to provide accurate judgments of themselves (Hofstee, 1994; Kolar, Funder, & Colvin, 1996).

As research continues to progress, it behooves investigators to begin to respond to the criterion problem instead of continuing to bypass it. The ultimate goal should be an appropriate, broad-based accuracy criterion, which would ideally derive from multiple modes of assessment including the target's self-perspective, ratings from knowledgeable informants, clinical judgments, behavioral measures, life outcome data, and perhaps even biological information (Funder, 1995). We propose that research that seeks to go beyond single operationalizations should refer to the outcome as *realistic accuracy*, in order to distinguish this construct from single operations, including self-other agreement. The current study moves towards the measurement of realistic accuracy with a criterion that is the average rating on each of 100 specific personality characteristics from judgments rendered by the self, two acquaintances from everyday life, and a clinical psychologist who conducted a life-history interview. Our operational definition of realistic accuracy will be the convergence between judgments and this broad-based criterion.

The Realistic Accuracy Model

The Realistic Accuracy Model (RAM; Funder, 1995, 1999) describes a 4-stage interpersonal and cognitive process that is required for accurate personality judgment. First, the target of the judgment must display cues or behaviors that are *relevant* to the characteristic being judged in such a way and in contexts that are *available* to the judge. Then, the judge must *detect* these cues and correctly *utilize* them to make a judgment. RAM assumes that the four stages combine in a multiplicative fashion, and if there is failure at any of the four stages then accuracy will not be possible.

Four factors can influence the accurate judgment process by affecting one or more stages of RAM (Funder, 1995): the good judge (Kolar, 1995; Taft, 1955; Vogt & Colvin, 2003), good target (Colvin, 1993), good trait (Borkenau & Leibler, 1993; Funder & Dobroth, 1987; John & Robins, 1993; Norman & Goldberg, 1966) and good information (Blackman & Funder, 1998; Funder & Colvin, 1988; Funder et al., 1995). The two aspects of the good information moderator, information quantity and information quality, are the focus of the current paper.

Information Quantity

A fair amount of research has examined the relationships between information quantity and two aspects of personality judgment, consensus and self-other agreement, and to our knowledge research has not previously been published on the relationship between information quantity and realistic accuracy. Past findings with consensus and self-other agreement have been mixed.

In cross-sectional designs, in which different participants are involved at each level of acquaintance, findings concerning consensus have both supported and not supported the acquaintanceship effect. Blackman and Funder (1998) experimentally manipulated level of acquaintance by having some people observe a target interacting for 5 or 10 minutes and other people observe a target for 25 or 30 minutes, and found that consensus did not increase with acquaintance. However, researchers have found that consensus is higher among real-life acquaintances than relative strangers (i. e. people who viewed dyadic interactions for either 5 minutes (Funder & Colvin, 1988) or 25 or 30 minutes (Blackman & Funder, 1998)). Furthermore, a meta-analysis suggests that for agreeableness, conscientiousness, neuroticism, and openness to experience, consensus is greater among long-term acquaintances than people who were unacquainted or had only interacted once (Kenny et al., 1994). In general, higher

levels of consensus have been associated with higher acquaintance in cross-sectional studies in which there was a relatively large difference in the amount of available information between the high and low acquaintance groups

Cross-sectional studies of self-other agreement have generally supported the acquaintanceship effect. For example, self-other agreement was higher among people who watched videotaped unstructured dyadic interactions for 25-30 minutes vs. 5-10 minutes (Blackman & Funder, 1998), people who had been roommates for more than 10 months vs. less than 10 months (Bernieri at al., 1994), people who indicated they knew the target of judgment extremely well vs. not at all (Paunonen, 1989), and real-life acquaintances vs. people who watched 5-minute unstructured dyadic interactions (Funder et al., 1995). In cross-sectional designs, information quantity appears to have a positive linear relationship with self-other agreement across all levels of acquaintance. However, a complication common to cross-sectional studies of consensus and self-other agreement is that real-life acquaintances have self-selected to know a person for longer, which introduces confounds such as similarity to the target and liking of the target.

In longitudinal research designs in which the same participants are involved at each level of acquaintance, previous findings suggest that as acquaintance increases, consensus generally remains constant while self-other agreement generally increases. Consensus has been shown to stay about the same as acquaintance increases across a variety of samples, situations, and lengths of time, including previously-unacquainted students who participated in 30-minute group interactions on four consecutive days (Park & Judd, 1989); students who participated in weekly 20-minute group meetings for seven weeks (Paulhus & Bruce, 1992); and students living near each other in a dorm from 2 weeks to 33 weeks after the start of a semester (Park, Kraus, &

Ryan, 1997). Furthermore, a meta-analysis suggested that consensus did not increase among people who interacted from between eight minutes and two hours in both laboratory and naturalistic settings (Kenny et al., 1994). On the other hand, an examination of self-other agreement provided evidence that self-other agreement generally increased among students who participated in weekly 20-minute group meetings for seven weeks (Paulhus & Bruce, 1992).

One possible explanation for the inconsistent findings concerning the acquaintanceship effect is that consensus was used as the outcome, which is not necessarily the appropriate way to approach this issue if one wishes to determine the extent to which judgments become more *accurate* with increased acquaintance. It is possible that the ratings of the judges in the studies mentioned previously did become more accurate across time, but this change might not be reflected in consensus if the ratings continued to have the same level of similarity with each other (Blackman & Funder, 1998). In fact, recent findings by Borkenau and colleagues (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004) suggest that acquaintance and accuracy are positively related when acquaintance is operationalized as the number of behavioral episodes rated by different judges that were included in a composite judgment. In the current project the relationships between information quantity and consensus and self-other agreement will be examined to allow for comparison of our results with past findings, and with current findings concerning realistic accuracy.

Another possible limitation of past findings is that researchers have simply assumed that people who have known each other longer have also acquired more information about each other, which is not necessarily true. One way to determine whether people in longer relationships have actually acquired more personality-relevant information is to directly measure the amount of objective information they know about the target. In the current study this goal

will be sought by asking participants a series of factual questions about their interaction partners that sample from the kinds of information people might learn in an initial interaction. We predict that people who interact for longer will acquire more of this information, concomitant to picking up more information relevant to personality, and therefore we expect to find a moderate positive relationship between information acquisition and realistic accuracy, consensus, and self-other agreement.

The informed reader will realize that the information quantity analyses amount to a conceptual replication of Blackman and Funder (1998), although the current project goes beyond Blackman and Funder's study and other research in several important ways. First, participants in Blackman and Funder's experiment were exposed to targets via videotaped two-person interactions, whereas participants in the current study actually interacted with each other (in three-person groups) before making personality judgments, which may have different implications for the way judges detect and use information when making personality judgments. More importantly, many researchers, including Blackman and Funder, have used self-other agreement as a proxy for accuracy, which, as discussed previously, has several limitations. The current study goes beyond self-other agreement and examines a construct closer to the ideal of realistic accuracy by using a broad-based accuracy criterion.

Information Quality

Although a considerable amount of research has focused on the relationship between information quantity and personality judgment, surprisingly little research has focused on an arguably even more important aspect of information – information quality, or the amount of *personality-relevant* information that becomes available within a given period of time. Intuitively, it seems apparent that not all kinds of information contribute equally to the process of

accurate judgment, as it is possible to be acquainted with someone for only a short time but to know him or her very well, or to be acquainted with someone for years and still know little about that person. From a research perspective, the first step in examining information quality is to determine the extent to which it is related to indicators of accuracy, including realistic accuracy, consensus, and self-other agreement, which is one goal of the current paper.

One property of a situation that might influence the quality of the information likely to be revealed in it is the situation's *strength* (Snyder & Ickes, 1985). *Strong* situations limit the range of behavior that people display because they include explicit rules or evoke implicit norms to which people generally adhere. At the other extreme, *weak* situations allow for considerable behavioral variation because there are few rules or norms for typical behavior. Therefore, in comparison to a strong situation, a weak situation should allow for more behavioral variation, the availability of higher quality individuating information, the acquisition of more objective information, and greater judgment accuracy.

Another possible dimension of information quality is the type of information that is exchanged. A casual interaction may include anything from trivial interchanges about the weather to intimate revelations of personal concerns. Anderson's (1984) ground-breaking study found that conversations about thoughts and feelings yielded more self-other agreement in personality ratings than did conversations about hobbies and activities. The present study is designed to begin the investigation of what occurs when people have the specific goal to get to know each other better by combining both aspects of information quality just discussed. A strong situation will be experimentally induced by giving participants the specific task of answering a long set of trivia questions, and two weaker situations will be represented by an unstructured conversation and by a setting in which participants are asked to get to know one

another as well as possible. We assume that more personality-relevant information will be available in the latter of the two weak situations, partly as the result of the type of information that is exchanged. RAM's prediction is that realistic accuracy, self-other agreement, and consensus will all increase along with information quality across these three conditions.

The Weighted-Average Model

The Weighted-Average Model (WAM; Kenny, 1991; 1994) offers a detailed description of the basis of consensus in personality judgment. The model predicts that consensus will not increase with acquaintanceship when judges see completely overlapping behaviors of the target and interpret what they see in exactly the same way (Kenny, 1991). However, when these assumptions are replaced by the more realistic expectations that different judges will interpret behaviors in different ways and will not detect exactly the same behaviors of the target, WAM predicts that consensus will increase rapidly at low levels of acquaintance, and remain about the same across higher levels of acquaintance. This reasoning leads us predict, on the basis of WAM, that information quantity will be positively related to consensus, but the difference will primarily be driven by an increase between low levels and medium to high levels of acquaintance, rather than between medium and high levels.

Research Questions

With the theoretical basis of both the Realistic Accuracy Model (Funder, 1995) and the Weighted-Average Model (Kenny, 1991) in mind, the current paper examines how quantity and quality of available information affect realistic accuracy, consensus, and self-other agreement.

Hypothesis 1: Participants who interact for longer periods of time will have access to more personality-relevant information and therefore achieve higher levels of realistic accuracy, consensus, and self-other agreement. To examine this issue, personality judgments following 3-

person unstructured interactions of three lengths (minimal interaction, 50-minutes, and three-hours) will be compared. Based on RAM (Funder, 1995, 1999), we expect that all three indicators of personality judgment - realistic accuracy, consensus, and self-other agreement - will increase linearly across these experimental conditions. Even though RAM is a model for accuracy, it can also be used to make predictions about consensus if one assumes that personality is something real, in which case two accurate judgments of personality must agree with each other, and therefore consensus can be expected to be high at sufficiently high levels of accuracy. Based on WAM (Kenny, 1991), we expect consensus to increase between the low and medium quantity conditions, and to stay about the same between the medium and high quantity conditions. Thus, the predicted relationships between information quantity and consensus differ slightly depending on which model is used, and both predictions will be tested.

Hypothesis 2: Participants who interact in situations in which more personality-relevant information is likely to be available will achieve higher levels of realistic accuracy, consensus, and self-other agreement. In one of the first experimental investigations to directly examine this issue, three 50-minute interactions that vary in the amount of personality-relevant information likely to be revealed in them will be compared. The low quality condition is a strong situation that allows for little behavioral variation, and is expected to elicit the least amount of personality-relevant information. The medium and high quality conditions are both relatively weak situations that are expected to elicit more behavioral variation and therefore higher levels of all indicators of accuracy than the strong situation. Between the two weak situations, participants who were given the objective to get to know each other are expected to share more personality-relevant information, such as thoughts and feelings (Anderson, 1984), and achieve higher levels of realistic accuracy, consensus, and self-other agreement than participants who were simply

instructed to talk about whatever they would like. Therefore, based on RAM, we predict a linear relationship between information quality and all three indictors of accuracy. We can also make a prediction based on WAM. Even though the model does not directly consider information quality, we have defined information quality as the availability of more personality-relevant information with time held constant, so that the acquaintanceship parameter of WAM should increase with information quality. Therefore, based on WAM, we again predict that consensus will increase more between the low and medium quality conditions and remain about the same between the medium and high quality conditions.

Methods

Participants

A total of 506 undergraduate students participated in the Riverside Accuracy Project – Phase 2 (RAP-II) and were paid $10 per hour for their time. A core group of 180 target participants (90 males and 90 females) were recruited via announcements made in psychology classes and fliers placed on bulletin boards advertising "Research on Personality." These participants were asked to recruit two close acquaintances (for a total of 326 acquaintance informants) to provide personality descriptions of themselves and the target participants. The ethnic breakdown of the target participants is 38% Asian-American, 20% Hispanic, 14% Caucasian, 12% African American, and 16% other or not specified. This is the second paper to come out of the extensive RAP-II data set, and the analyses do not overlap with the previous project, which focused on the constructs of ego-control and ego-resiliency (see Letzring, Block, & Funder, in press), or with future planned projects.

Overview of Procedures

Target participants came to the lab on three separate occasions and also completed three packets of self-report questionnaires outside of the lab, only some of which will be used in the current analyses. During the first session, participants interacted in three-person groups, in one of five experimental conditions, and then made personality judgments of their two interaction partners. During the third session, participants were interviewed by a clinical psychologist who described their personality after the conclusion of the interview. Finally, acquaintance informants completed questionnaires regarding the target participant with whom they were acquainted.

Measures

California Adult Q-set (CAQ; Block, 1961, as modified for use by nonprofessionals by Bem & Funder, 1978). The CAQ item set consists of 100 carefully-formulated descriptive statements about personality (e.g., is critical/skeptical, is personally charming, is cheerful). In the *Q-item rating*, each item is responded to using a 9-point Likert-type scale ranging from 1 (not at all characteristic) to 9 (highly characteristic). Participants rated themselves and their group interaction partners using the Q-item rating, and acquaintance informants rated the target participants by also using the Q-item rating. Participants rated both group interaction partners at the same time for each item of the CAQ, which implies that for each item the interaction partners were directly compared with each other at the time of the judgment. We expected this procedure to help the judge make individuating and thoughtful ratings for each item and both targets. Clinicians rated the target participants using the traditional *Q-sort* method, in which the 100 items are placed into a forced-choice, quasi-normal distribution so that each category (1 through 9) contains a pre-determined number of items.

Information Fact Sheet. An Information Fact Sheet was constructed for RAP-II that consists of open-ended questions that were designed to sample from the wide array of information that might become available during the experimental interactions. Participants attempted to answer questions about their interaction partners including their age, place of birth, political affiliation, future goals, family makeup, and so forth. No attempt was made to assess the information that participants picked up in any comprehensive fashion, or even to focus specifically on personality-relevant information (because techniques for either goal are presently unavailable). Rather, we aimed simply to obtain a small sample of the information that one person might learn about another during a casual interaction, with the assumption that scores on this measure would reflect the amount of information, including personality-relevant information, acquired overall. The Information Fact Sheet was completed by the target participants about their group interaction partners immediately following their ratings of personality, and was completed on a later occasion by each participant about himself or herself. Each question was scored correct (1 point) if the answer was equivalent to the answer given by the self, half (.5 points) if the answer was similar, but not equivalent, to the answer given by the self, or incorrect (0 points) if the answer was different from the answer given by the self or if there was no response. Each Information Fact Sheet was scored by two independent judges and any discrepancies in scoring were reconciled by a third judge. The reliability of the 19 items is high enough to justify using the sum as a score representing information acquisition (Cronbach's *alpha* = .77).

Procedures

Pre-session

At the pre-session participants were given the first take home packet of questionnaires and scheduled into one of the five experimental conditions. Each participant was shown a folder containing pictures of the other participants in the same group to determine whether he or she had ever seen the other participants. This procedure ensured that none of the participants had prior acquaintance with their interaction partners prior to the experimental group interaction.

Experimental Conditions

Target participants interacted in the lab in one of five experimental conditions designed to vary the quantity and quality of interpersonal information likely to become available in them. Each of the groups consisted of three previously unacquainted target participants in one of four possible gender compositions (i.e., all female (FFF), all male (MMM), two females and one male (FFM), and one female and two males (FMM)). An equal number of the four gender compositions were assigned to each experimental condition. A total sample size of 180[1] allows for a balanced design with three groups of each gender composition per condition, for a total of 12 groups per condition. Preliminary analyses determined that realistic accuracy and consensus were more similar than different across gender compositions, so all gender compositions will be analyzed together.

In all conditions, participants were seated in a room at a round table and given verbal instructions. At the conclusion of the interaction, participants rated their two partners using several measures including the rating form of the CAQ and the Information Fact Sheet.

1. Minimal Information Condition. Participants rated the other two participants in the room immediately following a short set of instructions explaining that, "we are studying first impressions," and "do the best that you can describing your first impressions of the other two people in the room." They were asked not to speak to each other, but were in each others'

physical presence for as long as it took to complete the ratings (less than one hour). This situation provides an empirical estimate of the baseline of realistic accuracy, consensus, and self-other agreement, and is used as the low-level condition for information quantity.

2. *Trivia-Quiz Condition.* Participants were presented with a packet containing 380 trivia-type questions, each of which had a single correct answer. They were informed that they would have 50 minutes to work jointly through the packet of questions and arrive unanimously at what they believed to be the correct answers. This context was designed to be a strong situation that leaves relatively little room for extraneous commentary or the disclosure of personality-relevant information, because the group would spend the entire interaction time responding to the trivia questions. This condition is used as the low-level condition for information quality.

3. *Short Unstructured Condition.* Participants were told that they could talk about anything they liked for the next 50 minutes. No attempt was made to direct or suggest what should be done over the course of the interaction. This context provides a weak situation in that participants are able to do and say whatever they would like and exhibit individual differences in social behavior, which should include at least some personality-relevant information on which judgments could be based. This condition is used as the mid-level condition for both information quantity and quality.

4. *Long Unstructured Condition.* Participants were told that they could talk about anything they would like for the next three hours, and as in the Short Unstructured condition no attempt was made to direct or suggest what should be done over the course of the interaction. The experimenter returned to the room halfway through the interaction to give participants a short break, at which time snacks were provided and a restroom break was offered.   As in the Short Unstructured condition, this context provides a weak situation with very little structure and

should include at least some personality relevant information on which judgments could be based. This condition is used as the high-level condition for information quantity.

     5. Get to Know (GTK) Condition. Participants were told that their "task for the next 50 minutes is to get to know each other as well as possible and to learn as much as you can about what type of person each of you is."   This context provides a weak situation that is expected to yield a wide range of behaviors relevant to personality, and the objective of getting to know each other is expected to guide participants to probe for and to remember information related to personality. This condition is used as the high-level condition for information quality.

Life-History Interview

     Participants were individually interviewed by one of four professionally trained (i.e. MSW, MA in counseling, or PhD in clinical psychology) and licensed clinical psychologists who had experience with a college population. With consent of the participants, all interviews were videotaped. The clinicians conducted a one-hour semi-structured life history interview adapted from a protocol used for many years by the Institute of Personality Assessment and Research (IPAR; Craik et al., 2002). The protocol used in the current study was adapted to better apply to college students and sought to capture a broad range of personality-relevant information without explicitly asking about sensitive topics and risky behaviors. Each interview started with the clinician asking the participant to "tell me something about yourself," and then covered a broad range of topics including college and academic experiences, future plans, interpersonal relationships, and childhood and family history. In conclusion, each participant was asked to describe "a defining event in your life that had a significant impact on or changed your life in some way."[2]   Following the interview, the clinician completed a Q-sort description of the target participant. Some of the interviews were later viewed by a second clinician, and in these cases

the second clinician also described the participant's personality using the CAQ.[3] When two

ratings of the participant were available, a composite score was computed for each item and used

in subsequent analyses. The average inter-clinician profile agreement is $r = .50$ ($SD = .17$)[4] and

the coefficient of internal consistency based on a composite of the two raters is .67.

Acquaintance-Informant Ratings

Participants were asked to provide the names and contact information of the two people

who knew them best at the university. Acquaintances came into the lab and provided

descriptions of the target participant by whom they had been identified, using the rating format

of the CAQ. The average profile agreement for the 100 items as rated by two acquaintances is r

$= .40$ ($SD = .19$).[5] The average coefficient of internal consistency, or the dependability of the

profile based on a composite of the two acquaintances, is .57.

Results

The focus of our analyses will be on realistic accuracy, as this project is one of the first to

explore the construct, and the first to our knowledge to examine the relationships between

realistic accuracy and experimentally-manipulated information quantity and quality. We will

also provide evidence concerning our predictions for consensus and self-other agreement.

Finally, we will determine whether participants in longer and higher quality interactions did

indeed acquire more information (based on a sample of information possibly available in such

interactions), and the degree to which information acquisition is related to realistic accuracy,

consensus, and self-other agreement.

There are several possible ways to analyze these data. We will use the traditional social

psychological approach that focuses on the group level. This means that we computed average

scores for realistic accuracy, consensus, self-other agreement, and information acquisition *within*

each group and used these numbers for contrast analyses and correlational analyses. Our predictions are concerned with properties of the context, not of the individual, and a group-level analysis reflects this orientation. While none of the basic conclusions of this study changes when data are examined at different levels of analysis, an advantage of analyzing data at the level of the group is that it allows us to perform all analyses in a consistent manner.

Analyses

Realistic accuracy criterion

Many researchers have used self-other agreement as a proxy for accuracy, but in the current project we strive to go beyond this simple operationalization by comparing judgments to a broader-based criterion composed of ratings by three types of knowledgeable informants: the self, two acquaintances nominated by the target, and a professionally trained clinician-interviewer. The 100 items of the CAQ were rated by each type of rater, and for each target an average for each item was then computed across the multiple types of informants. The mean alpha reliability was high enough to justify the use of these average ratings (*mean* = .42, range = .08 - .72)[6]. The similarity, in the form of a correlation coefficient, between ratings by a participant following the experimental group interaction and the accuracy criterion yields a realistic accuracy score.

Profile analysis

Profile analysis allows for an examination of judgments in regards to a target's overall personality by utilizing the entire set of 100 CAQ items in a single analysis. The computation of a profile correlation simply involves correlating two sets of ratings of the same target, across the 100 items of the CAQ. First, to examine realistic accuracy, the ratings of each interaction partner were correlated with the accuracy composite, which yields six scores for each group (2

scores for each of 3 judges), and these scores can be Z-transformed and averaged so that each group has a single realistic accuracy score (see Figure 1). Next, to examine consensus, the ratings provided by the two interaction partners of each target were correlated with each other, resulting in a profile correlation that represents the degree of similarity in the partner's ratings of that target. This analysis yields three consensus scores for each group, which can be Z-transformed and averaged so that each group has one consensus score. Finally, to examine self-other agreement, the ratings of each interaction partner were correlated with the self-ratings. As in realistic accuracy, this procedure yields six self-other agreement scores per group, and these scores can be Z-transformed and averaged so that each group has a single self-other agreement score. To determine the mean level for each type of judgment in each type of interaction, the Z-transformed correlations for all groups in each experimental interaction were averaged.

Cronbach (1955) was one of the first to point out that profile correlations may be confounded with several other elements besides differential accuracy, including elevation, differential elevation, and stereotype accuracy, that result from the use of response sets or a reliance on stereotypes. The effects of response sets should not be of concern for between-group comparisons in the current experiment because participants were randomly assigned to one of five experimental conditions, and therefore participants with various response sets should be equally distributed among them, and stereotype accuracy should have equivalent effects on the magnitude of the profile correlations across conditions. For this reason, we use the raw profile correlations in our cross-group comparisons, as did Blackman and Funder (1998) when they conducted similar analyses.

However, it is true that as a result of stereotype accuracy, the baseline correlation between two sets of ratings is not expected to be zero because some items of the CAQ are

generally rated higher than others, regardless of the target (Blackman & Funder, 1998). In order

to determine baseline profile correlations within the current data, the correlations among 30

random pairs of ratings were computed. For realistic accuracy, the profile correlations were

computed between the ratings of one target and the accuracy composite for a different target,

which resulted in a baseline profile correlation of $r = .36$. For consensus, the profile correlations

were computed between two interaction partners of different targets, which resulted in a baseline

profile correlation of $r = .25$. For self-other agreement, the profile correlations were computed

between the ratings of one target and the self-ratings of a different target, which resulted in a

baseline profile correlation of $r = .28$. These baseline correlations may at first seem somewhat

high, but when one realizes that all participants in the sample are college students, and that in

general college students are similar to each other in many ways, it makes sense that even random

pairings result in profile correlations of moderate magnitude. Profile correlations greater than

these baseline correlations indicate differential accuracy (Blackman & Funder, 1998).

Information Quantity

To test Hypothesis 1, we examined personality judgments at three levels of information

quantity: low (Minimal Information/no interaction), medium (Short Unstructured/50-minute

interaction), and high (Long Unstructured/3-hour interaction). First, profile correlations were

computed for each group within each experimental condition, as described in the Analyses

section. The profile correlations were transformed using Fisher's r to Z transformation, and then

contrast analysis (Rosenthal, Rosnow, & Rubin, 2000) was used to determine the degree to

which the profile correlations followed predictions based on RAM and WAM. Contrast analysis

tests focused questions by assigning weights to experimental conditions that reflect a prediction

concerning the structure of the data. Based on RAM, we predicted that there would be a linear

relationship between information quantity and realistic accuracy, consensus, and self-other agreement, for which the appropriate contrast weights are -1 (Minimal Information), 0 (Short Unstructured), and +1 (Long Unstructured). Based on WAM, we predicted a slightly different relationship between information quantity and consensus in which there is an increase at low levels of acquaintance and then a leveling off, for which the appropriate contrast weights are -2 (Minimal Information), +1 (Short Unstructured), and +1 (Long Unstructured).

As predicted, we found that realistic accuracy had a positive linear relationship with information quantity (Minimal Information mean $r = .29$, Short Unstructured mean $r = .45$, Long Unstructured mean $r = .46$; $r_{contrast} = .52$, $F_{contrast} (1, 33) = 12.28$, $p < .005$; see Figure 2). Information quantity also had a positive linear relationship with consensus (Minimal Information mean $r = .25$, Short Unstructured mean $r = .41$, Long Unstructured mean $r = .44$; $r_{contrast} = .44$, $F_{contrast} (1, 33) = 8.12$, $p < .01$), and self-other agreement (Minimal Information mean $r = .21$, Short Unstructured mean $r = .36$, Long Unstructured mean $r = .39$; $r_{contrast} = .49$, $F_{contrast} (1, 33) = 10.55$, $p < .005$). The consensus data are also in line with the prediction based on WAM ($r_{contrast} = .48$, $F_{contrast} (1, 33) = 9.24$, $p < .005$).

Note that for all indicators, the mean profile correlations for the Minimal Information condition are approximately equal to or smaller than the random baseline correlations of .36 for realistic accuracy, .25 for consensus, and .28 for self-other agreement, but the mean profile correlations of the Short Unstructured and Long Unstructured conditions are all higher than the random baseline correlations. This pattern suggests that even though realistic accuracy, consensus, and self-other agreement can reach moderate strengths without any interaction, this strength is at or below chance levels. It is only when people are allowed to interact that differential realistic accuracy, consensus, and self-other agreement are achieved.

Information Quality

Recall that information quality refers to how relevant information is to personality. To test our predictions concerning information quality, participants interacted in three experimental conditions in which information quantity was held constant while information quality was manipulated. We again used contrast weights reflecting the prediction of a linear relationship between information quality and realistic accuracy, consensus, and self-other agreement, with weights of –1 (Trivia), 0 (Short Unstructured), and +1 (GTK) to test the prediction based on RAM, and contrast weights of -2 (Trivia), +1 (Short Unstructured), and +1 (GTK) to test the prediction for consensus based on WAM. Contrast analyses were performed on the Z-transformed correlation coefficients.

As predicted, it was found that realistic accuracy had a positive linear relationship with information quality (Trivia mean $r = .36$, Short Unstructured mean $r = .45$, GTK mean $r = .49$; $r_{contrast} = .43$, $F_{contrast} (1, 33) = 7.34$, $p < .02$; see Figure 3). Information quality also had a positive linear relationship with consensus (Trivia mean $r = .29$, Short Unstructured mean $r = .41$, GTK mean $r = .41$; $r_{contrast} = .33$, $F_{contrast} (1, 33) = 3.94$, $p < .10$), and self-other agreement (Trivia mean $r = .27$, Short Unstructured mean $r = .36$, GTK mean $r = .39$; $r_{contrast} = .38$, $F_{contrast} (1, 33) = 5.70$, $p < .02$). The consensus data are also in line with the prediction based on WAM ($r_{contrast} = .37$, $F_{contrast} (1, 33) = 5.32$, $p < .05$).

Note that again, for all three indicators the mean profile correlations for the lowest quality condition (Trivia) are approximately equal to the random baseline correlations, but the mean profile correlations of the Short Unstructured and GTK conditions are all higher than the random baseline correlations. This pattern suggests that not just any context of observation will afford differential judgment, but that differential realistic accuracy, consensus, and self-other

agreement are only achieved in the conditions that allow for the exchange of moderate to high

quality information.

Information Acquisition

In order to determine whether participants in interactions of greater quantity and quality

actually acquired more information about each other, we compared the scores that were achieved

on the Information Fact Sheet across experimental conditions. It was found that judges answered

more questions correctly when they interacted for longer periods of time (Minimal Information

mean = 2.80, Short Unstructured *mean* = 5.17, Long Unstructured *mean* = 7.00; $r_{contrast}$ = .81,

$F_{contrast}$ (1, 33) = 64.39, p < .001; see Figure 4), and in higher quality situations (Trivia *mean* =

1.65, Short Unstructured *mean* = 5.17, GTK *mean* = 5.87; $r_{contrast}$ = .82, $F_{contrast}$ (1, 33) = 67.53,

p < .001; see Figure 5). These findings concerning a small sample of the available information

imply that more information overall was acquired by participants in longer and higher quality

interactions, where judgment was also more accurate.

We can also examine the degree to which information acquisition is related to realistic

accuracy, self-other agreement, and consensus across all groups. RAM predicts that people who

acquire more information that is relevant to personality will make more accurate judgments of

personality, and it is reasonable to assume that participants who acquired more of the sample of

information asked about on the Information Fact Sheet also acquired more personality-relevant

information, broadly speaking. If this is true, then there should be a positive relationship

between scores on the Information Fact Sheet and realistic accuracy, self-other agreement, and

consensus. This prediction can be tested by correlating the average Information Fact Sheet score

of each group with the average realistic accuracy, consensus, and self-other agreement scores of

each group, across all experimental conditions. As predicted, moderate positive relationships

were found between information acquisition and realistic accuracy ($r = .31$, $p = .02$), self-other agreement ($r = .30$, $p = .02$), and consensus ($r = .30$, $p = .02$), indicating that in contexts where people acquired more objective information that was sampled with the Information Fact Sheet, they also judged their partners with higher average degrees of accuracy, consensus, and self-other agreement.

Discussion

The preceding analyses provide tests of the relationships between both aspects of good information and the accuracy of personality judgment. Information quantity, or acquaintance, was positively related to realistic accuracy, consensus, and self-other agreement. Furthermore, information quality was manipulated while information quantity was held constant, and support was found for the prediction that realistic accuracy, consensus, and self-other agreement are higher in situations in which personality-relevant information is more likely to be available.

Information Quantity

Both RAM and WAM predict that accuracy will increase in relation to information quantity, but previous research has only tested this prediction using self-other agreement, which is a limited form of accuracy. The current research confirms past findings, and provides one more piece of evidence for the positive relationship between level of acquaintance and self-other agreement. In addition to confirming past findings, the current research takes an additional step in the examination of the acquaintanceship effect by using a broad-based accuracy criterion based on the ratings of several people who knew the target well. Such a criterion allows for a more precise test of the acquaintanceship effect than was possible in past work. In line with our prediction, realistic accuracy was positively related to information quantity.

Both RAM and WAM can be used to make predictions concerning consensus. RAM predicts that realistic accuracy will increase across all levels of acquaintance, and because people who are highly accurate about something real must also be in agreement, RAM predicts that consensus will also increase with acquaintance (Funder, 1995). WAM is specifically aimed at explaining consensus, and the prediction is that consensus will increase at low levels of acquaintance and then will stay about the same even as acquaintance continues to increase (Kenny, 1991). These predictions are fairly similar, and we found support for both predictions. Furthermore, the results of the contrast analyses were not different enough to suggest that one prediction is superior to the other. Overall, these findings support the existence of the acquaintanceship effect and the role of information quantity in accurate judgment, and our interpretation was bolstered by the further finding that participants learned more facts about each other in the longer interaction conditions.

Information Quality

To our knowledge, the current paper provides the first direct evidence of the positive relationships between information quality and realistic accuracy, consensus, and self-other agreement. Information quality was manipulated by varying situational strength (Snyder & Ickes, 1985) and instructions providing an objective for the interaction. For the low-quality interaction, participants interacted in a strong situation with specific instructions to complete a packet of trivia-type questions. We expected participants in this condition to exhibit little behavioral variation and to not share information about thoughts and feelings, and therefore only limited amounts of personality-relevant information would become available. For the medium- and high-quality interactions, participants interacted in relatively unstructured situations that allowed for a good deal of behavioral variation and larger amounts of personality-relevant

information to become available. The difference in these two conditions was the instructions informing the participants about the objective of the interaction. Participants in the medium-quality interaction were not provided with an objective, while participants in the high-quality interaction were given the specific objective of getting to know each other as well as possible. Participants in this latter condition were expected to reveal the most personality-relevant information and to pay the best attention to the interaction partners in an attempt to get to know them well, which would increase accuracy by affecting the availability and detection stages of RAM. Results confirmed these expectations with realistic accuracy, consensus and self-other agreement all increasing, along with factual knowledge, in the predicted order across the three conditions of information quality. However, the present data do not allow us to determine whether these effects should be attributed to what cognitive psychologists would call the *encoding* stage or the *decoding* stage. It may be that more relevant information is available to be detected (encoding), that the detected information is remembered and utilized better (decoding), or both.

Limitations and Future Directions

As in any study, our experimental design has limitations that may limit interpretation. First, the construct of realistic accuracy is a philosophical ideal, and any measurement of it must necessarily be an approximation. We have attempted to approach realistic accuracy by having several knowledgeable people, including expert judges, describe the personality of the target participants, which is not likely to tell us *exactly* what the target is like but is likely to bring us closer to reality than would ratings provided by any single informant, even the self. In future research we would hope to see criteria for realistic accuracy expanded even further to include direct behavioral measurements, important life outcomes, and perhaps even biological markers.

The present results involving realistic accuracy, consensus and self-other agreement were in general quite similar to each other. This finding could be taken as suggesting that realistic accuracy cannot be meaningfully distinguished from the simpler and more traditional operationalizations, or that constructing multi-faceted criteria is not worth the time and effort. We would caution against both conclusions. First, realistic accuracy is a hypothetical construct that is in principle quite different from any specific criterion for or measurement of it, and on theoretical grounds it is important to be clear about the distinction. Second, it is easy to imagine circumstances where self-other agreement, consensus, and realistic accuracy would diverge, such as when an absentminded job applicant tries to appear conscientious, and at our present state of knowledge it is wise to be alert to this possibility, and to develop tools to detect it when it happens.

Even though the contrast analyses suggest that there are moderate relationships between information quantity and quality and indicators of accuracy, this effect was largely driven by a differences between the low conditions (Minimal Information and Trivia) and the medium and high conditions (Short Unstructured, GTK, and Long Unstructured), with the medium and high conditions being only slightly different, albeit consistently in the predicted direction. The pattern of findings would have been even more convincing had there been a larger difference in realistic accuracy between the medium and high-level conditions. We propose that the present evidence supports the role of quantity and quality in accuracy, but that our highest conditions may not have increased acquaintance enough or provided enough motivation to show a substantial increase in indicators of accuracy beyond the medium levels. Future research should experiment further with techniques for providing participants large amounts of personality-relevant information of high quality.

Another reasonable next step would be to increase the external validity of our conclusions through a project in which interactions take place in real social contexts (vs. in the lab, as was the case in the current experiment). There are many situations in which previously-unacquainted groups of people interact and get to know each other over a period of time, such as first-year freshmen in college dorms, participants in volunteer corps who live and/or work together (i.e. Peace Corps, Jesuit Volunteer Corps), and elderly people who move into retirement communities. Information quantity could be examined in such situations by obtaining judgments about others in the judge's living community at several stages of acquaintance. Information quality could be examined by obtaining information about the quality of interactions from both the target and the judge, perhaps by asking for reports of types of topics most often discussed and amount of time spent talking versus engaging in other activities. Such a project is likely to be informative about how information quantity and quality are related to realistic accuracy in situations and contexts outside of the laboratory, as well as to how these two aspects of information are related to each other and personality judgment.

A third direction for future research is to determine the actual behavioral events during an interaction that are related to the achieved levels of realistic accuracy, self-other agreement, and consensus. Behavioral coding of the experimental interactions might increase our understanding of why information quantity and quality are related to accuracy, self-other agreement, and consensus. Such coding is in progress in our laboratory, but the process is extremely time-consuming and the results must await a future report.

Conclusion

Judgments of personality are more likely to achieve higher levels of realistic accuracy, consensus, and self-other agreement when judges have interacted with targets for longer periods

of time or in situations that allow or encourage revealing personality-relevant information.

These findings support several aspects of the Realistic Accuracy Model, including the

implications of information for the relevance, availability, and detection stages that make

accurate judgment possible, along with specific predictions of the Weighted Average Model

concerning consensus.

References

Allport, G. W. (1937). The ability to judge people. *Personality: A Psychological Interpretation.* (pp 499-522). New York: Holt.

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.

Anderson, S. M. (1984). Self-knowledge and social inference: II. The diagnosticity of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology, 46,* 294-307.

Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review, 85,* 485-501.

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.

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.

Block, J. (1961). *The Q-Sort Method in Personality Assessment and Psychiatric Research*. Palo Alto, CA: Consulting Psychologists Press.

Borkenau, P., & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality, 61,* 477-496.

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.

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

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

Craik, K. H., Ware, A. P., Kamp, J., O'Reilly, C. III, Staw, B., & Zedeck, S. (2002). Explorations of construct validity in a combined managerial and personality assessment programme. *Journal of Occupational and Organizational Psychology, 75*, 171-193.

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

Funder, D. C. (1980). On seeing ourselves as others see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality, 48,* 473-493.

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

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., & 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. (1993). Consensus, self-other agreement, and accuracy in

    personality judgment: An introduction. *Journal of Personality, 61,* 457-476.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of*

    Personality, 8, 149-162.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits:

    The big five domains, observability, evaluativeness, and the unique perspective of the

    self. *Journal of Personality, 41,* 521-551.

Jones, E. E. (1985). Major developments in social psychology during the past five decades. In G.

    Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 47-107). New

    York: Random House.

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.* New York: Guilford Press.

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.

Kolar, D. W. (1995). *Individual differences in the ability to accurately judge the personality*

    characteristics of others. Unpublished doctoral dissertation, University of California,

    Riverside.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality

    judgments by the self and knowledgeable others. *Journal of Personality, 64,* 311-337.

Letzring, T. D., Block, J., & Funder, D. C. (in press). Ego-control and ego-resiliency: Generalization of self-report scales based on personality descriptions from acquaintances, clinicians, and the self. *Journal of Research in Personality.*

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

Park, B., & Judd, J. 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.

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

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach.* New York, NY: Cambridge University Press.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology: Third edition.* Vol. 2 (pp. 883-947). New York: Random House.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52,* 1-23.

Vernon, P. E. (1933). Some characteristics of the good judge of personality. *Journal of Social*

    Psychology, 4, 42-57.

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality

    judgments. *Journal of Personality, 71,* 267-295.

Author Note

Tera D. Letzring, Department of Psychology, University of California, Riverside; Shannon M. Wells, Sanford Systems; David C. Funder, Department of Psychology, University of California, Riverside.

Correspondence concerning this article should be addressed to Tera D. Letzring, Department of Psychology, University of California, Riverside, Riverside, California, 92521. E-mail: hardit01@ucr.edu.

Footnotes

[1]Upon completion of data collection, it was found that two target participants were inadvertently included in the study twice. For both participants, the second condition that they were in was different from the first. Analyses were done with these groups included and excluded. The differences between these two sets of findings were within rounding error. In order to maintain a fully balanced design, results are reported from the data of all 60 groups.

[2]The full protocol for the clinical interview is available on the Riverside Accuracy Project's website at http://www.faculty.ucr.edu/%7Efunder/lab/publications.htm.

[3] When the participant scheduled for a live interview was unable to keep the appointment, clinicians were given the option to observe and rate a previously recorded interview. In this manner, 47 interviews were observed and rated by a second clinician.

[4]This number can be compared to 30 random pairings of ratings by clinicians who described different targets, which resulted in an average profile correlation of .29.

[5]This number can be compared to 30 random pairings of ratings by acquaintances who described different targets, which resulted in an average profile correlation of .24.

[6] Based on the correlations among types of raters, the accuracy criterion is appropriate to use for computing profile correlations (self-acquaintance $mean = .48$, $sd = .17$; self-clinician $mean = .39$, $sd = .14$; acquaintance-clinician $mean = .37$, $sd = .19$; Spearman-Brown for 3 raters $= .68$).

Figure Captions

Figure 1. Computation of profile realistic accuracy scores.

Figure 2. Mean realistic accuracy scores for experimental conditions used in the information quantity analyses.

Figure 3. Mean realistic accuracy scores for experimental conditions used in the information quality analyses.

Figure 4. Mean information acquisition scores for experimental conditions used in the information quantity analyses.

Figure 5. Mean information acquisition scores for experimental conditions used in the information quality analyses.

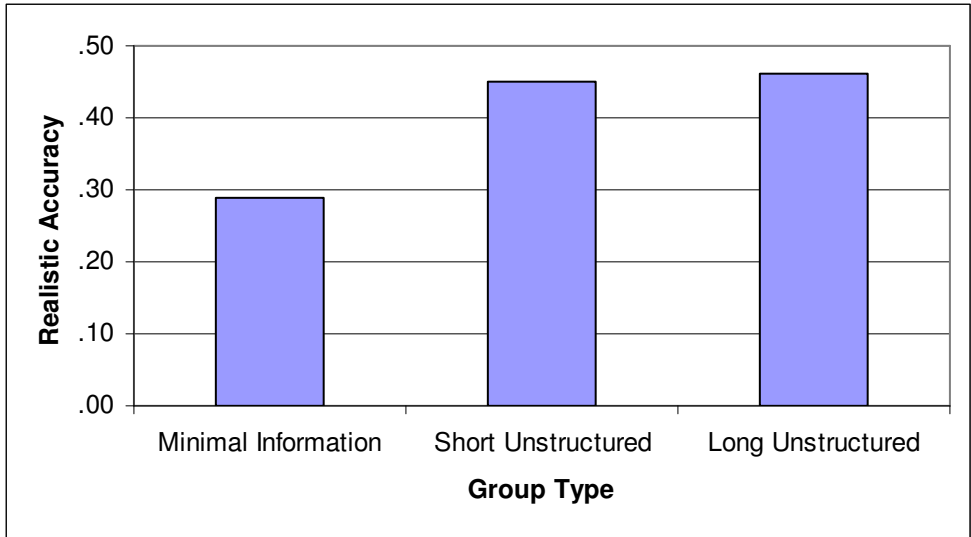|	Rating of partner A	Accuracy Criterion	Rating of partner B	Accuracy Criterion
CAQ 1	3	5	9	7
CAQ 2	7	8	5	6
...				
CAQ 100	1	3	4	8

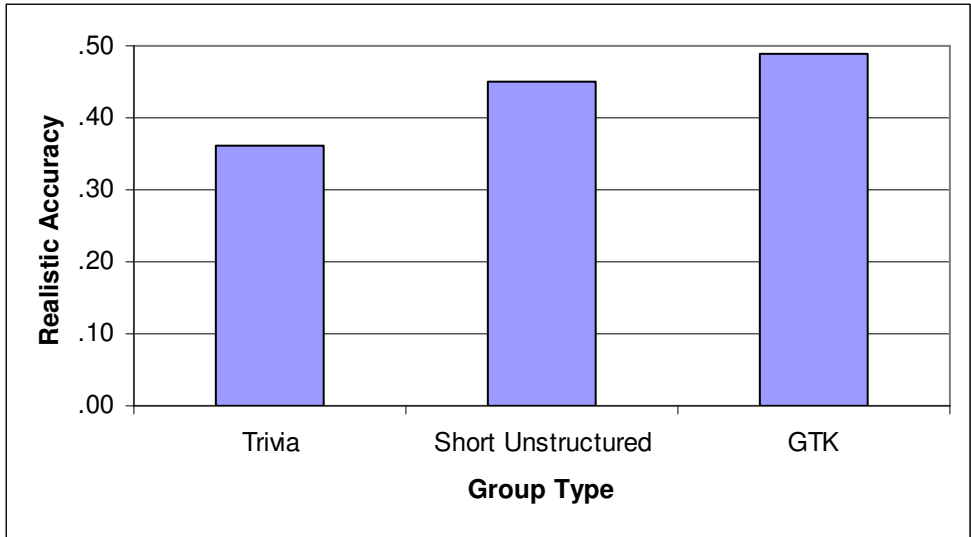Correlation for Partner A   Correlation for Partner B

Accuracy score of judge

Note. Profile correlations were computed for each of the 3 judges in each experimental group, and these scores were *Z*-transformed and averaged to compute scores for each group; the scores of all 12 groups in each type of experimental interaction were averaged to determine the accuracy for the corresponding interaction type.

Note. The random baseline profile correlation is $r = .36$.

Note. The random baseline profile correlation is *r* = .36.