Information Quantity and Quality Affect the Realistic Accuracy of Personality Judgment

Tera D. Letzring
University of California, Riverside

Shannon M. Wells
Sanford Systems, Incorporated

David C. Funder
University of California, Riverside

Triads of unacquainted college students interacted in 1 of 5 experimental conditions that manipulated information quantity (amount of information) and information quality (relevance of information to personality), and they then made judgments of each others’ personalities. To determine accuracy, the authors compared the ratings of each judge to a broad-based accuracy criterion composed of personality ratings from 3 types of knowledgeable informants (the self, real-life acquaintances, and clinician-interviewers). Results supported the hypothesis that information quantity and quality would be positively related to objective knowledge about the targets and realistic accuracy. Interjudge consensus and self–other agreement followed a similar pattern. These findings are consistent with expectations based on models of the process of accurate judgment (D. C. Funder, 1995, 1999) and consensus (D. A. Kenny, 1994).

Keywords: personality judgment, judgment accuracy, realistic accuracy, information quantity, information quality

Every day, people make personality judgments of other people. The person making the judgments (the judge) has known some targets for several years and has seen them in many different contexts (e.g., a sibling or a spouse), whereas other targets are people the judge has just met and has seen only in a relatively limited context (e.g., a new coworker, a bank teller). Not surprisingly, these judgments vary in how accurate they are, perhaps because they are made on the basis of varying degrees of acquaintance and information that vary in how relevant they are to personality.

These two aspects of the information available to a judge are referred to as information quantity, the sheer amount of information that is available, and information quality, the degree to which the available information is relevant to personality (Blackman & Funder, 1998; Funder, 1999). The concept of information quantity is straightforward in that it assumes that judges who have access to more information about the personality of the target will make more accurate judgments of personality. The effect of information quantity has also been referred to as the acquaintanceship effect (Colvin & Funder, 1991; Kenny, 1991), because as people are acquainted longer they are assumed to acquire more information about each other. The concept of information quality is more complex, but at its core refers to the likelihood that even when information quantity is held constant, different contexts of acquaintanceship might vary in the degree to which personality-relevant information becomes available in them. For example, a context in which everybody acts the same would yield only low-quality information for personality judgment because information about individual differences in personality would not be revealed, whereas a context in which behavior is relatively free to vary would yield higher quality information because people would be more likely to behave in a way that reflects their personalities (Snyder & Ickes, 1985). It is also possible that different kinds of information—for example, thoughts and feelings as opposed to hobbies and activities (Anderson, 1984)—are differentially informative about personality. When more personality-relevant information is available, the accuracy of judgments of individual differences in personality is expected to be higher. Information quantity and quality are two aspects of good information, one of four factors proposed to be related to judgment accuracy (Funder, 1999).

It is difficult to examine information quantity and quality independently of each other in natural settings of acquaintanceship, because people are likely to emit a broader range of cues to personality and share more personality-relevant information with acquaintances they have known longer. For example, according to social penetration theory (Altman & Taylor, 1973), people share information about a larger number of the aspects of their person-
ality and share more personal information within these aspects as a relationship progresses, which suggests that information increases in both quantity and quality along with acquaintance.

In the current study, the quantity and quality of the information available to judges were experimentally and independently manipulated so that the effects of these two factors on the accuracy of personality judgment could be examined, in order to achieve two goals. The first goal was to further examine the effects of information quantity on judgment accuracy by (a) testing for the acquaintanceship effect in a new context and across different levels of information quantity and by (b) examining the effects of information quantity on a relatively new construct, realistic accuracy, in addition to the more traditional constructs of consensus and self–other agreement. The second goal of the present research was to directly examine the effects of experimentally manipulated information quantity on realistic accuracy, consensus, and self–other agreement, which is a more distinctive contribution, as this is one of the first empirical studies to examine these effects.

Indicators of Judgment Accuracy

Three indicators of judgment accuracy were assessed: realistic accuracy, consensus, and self–other agreement.

Realistic Accuracy

Realistic accuracy refers to a hypothetical construct representing the level of agreement between a personality judgment and what a target is really like (Funder, 1995). This construct cannot be directly measured by any single personality or behavior rating, as any single rating is highly uncertain as an indicator of what a person is really like. Instead, the ideal of realistic accuracy can be approached to the degree that multiple methods of measurement are used and combined to form a broad-based accuracy criterion for each target of judgment. The assumption is that this broad-based criterion is closer to what the target is really like than is any single rating, because random errors in the ratings should cancel each other out as more ratings are combined. In the current study, we formed an accuracy criterion by using a composite rating based on judgments from real-life acquaintances, psychologists who interviewed the targets, and the self.

The difficulty of the criterion problem has led many researchers to shy away from studying the accuracy of judgments of real people and to instead turn their attention to models of the cognitive processing of artificial stimuli (Funder, 1987, 1995). As research progresses, 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 such as hormone levels and functional MRI images (Funder, 1995). We propose that research that seeks to go beyond single operationalizations of the accuracy of personality judgment should refer to the outcome as realistic accuracy in order to distinguish this construct from other constructs that measure some other singular form of agreement or accuracy in personality judgment.

Consensus

Consensus can be measured more directly than realistic accuracy and is more commonly studied. Consensus is the level of agreement between the personality judgments rendered by two or more people about another person. Consensus can be used as an indicator of accuracy, but it is important to keep in mind that consensus is not necessarily related to what the person is actually like, because many people could agree about the personality of someone else, and all could be wrong, as is the case when inaccurate stereotypes are used to make judgments.

Self–Other Agreement

Self–other agreement can also be directly measured and has been used frequently in previous research (Ambady, Hallahan, & Rosenthal, 1995; Blackman & Funder, 1998). It refers to the level of agreement between judgments made about another person and that person’s self-judgments. The self-judgments are often implicitly assumed to reflect reality, and the judgments made by others are assumed to be accurate to the degree that they are related to the self-judgments. However, self–other agreement does not guarantee highly accurate judgments, because the self-rating will not adequately represent one’s real personality if people are unwilling or unable to provide accurate judgments of themselves (Hofstee, 1994; Kolar, Funder, & Colvin, 1996). Despite this complication of using self–other agreement to measure accuracy, many investigations refer to self–other agreement simply as accuracy (Ambady et al., 1995; Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Blackman & Funder, 1998). However, in the current article the term accuracy is used more specifically to describe realistic accuracy so that realistic accuracy can be differentiated from consensus and self–other agreement.

We report findings concerning all three indicators of judgment accuracy for two reasons. First, realistic accuracy, although related to consensus and self–other agreement, is far from synonymous with either of these operationalizations and could potentially yield different results. Second, consensus and self–other agreement have been used in many prior studies (e.g., Anderson, 1984; Blackman & Funder, 1998; Kenny, Albright, Malloy, & Kashy, 1994), and our presentation of these constructs allows comparison with the existing literature.

Methods of Personality Judgment

Realistic Accuracy Model

The Realistic Accuracy Model (RAM; Funder, 1995, 1999) describes an interpersonal and cognitive process that results in accurate personality judgment when all four of its stages are successfully completed. 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 use them to make a judgment. RAM assumes that the four stages combine in a multiplicative fashion so that a failure at any of the four stages will make accuracy impossible. For example, cues can only be detected if they are available.

Four factors are thought to influence the judgment process by affecting one or more stages of RAM (Funder, 1995): the good
judge (Allport, 1937; Kolar, 1995; Taft, 1955; Vernon, 1933; Vogt & Colvin, 2003), good target (Colvin, 1993), good trait (Borkenau & Liebler, 1993; Funder & Dobroth, 1987; John & Robins, 1993; Norman & Goldberg, 1966), and good information (Blackman & Funder, 1998; Funder & Colvin, 1988; Funder, Kolar, & Blackman, 1995). Two aspects of good information, information quantity and information quality, are the focus of the current article.

A prediction regarding the relations between information quantity and quality and all three indicators of accuracy can be generated on the basis of RAM. 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. A linear increase between information and realistic accuracy, consensus, and self–other agreement was used as the prediction based on RAM. This is a conservative prediction because it is not possible to determine the degree to which the manipulated levels of information quantity and quality are equally spaced and, therefore, whether the increases in accuracy should also be equally spaced.

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 led us to predict that consensus would increase between low and medium levels of information quantity and quality but would be the same between medium and high levels of information quantity and quality.

The prediction based on WAM was tested only for consensus because WAM is explicitly formulated as a model of the process of consensus, and it has been suggested that the relations between different indicators of accuracy and acquaintanceship are not necessarily the same (Kenny, 1991). For example, Kenny (1991) pointed out that accuracy can increase with acquaintance even as consensus stays the same, as is the case when all judges become more accurate but continue to agree with each other to the same degree.

It was not feasible to determine whether the prediction based on RAM was superior to the prediction based on WAM, as the two predictions were highly correlated. However, it was still interesting to determine how well the obtained data fit each prediction.

Past Findings and Rationale for Current Experiment

Information quantity has received a fair amount of research attention, and although the findings have been mixed, 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 (e.g., Anderson, 1984).

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. To our knowledge, research has not previously been published on the relationship between information quantity and realistic accuracy. Past findings on consensus and self–other agreement have been mixed.

In between-subjects 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 min and other people observe a target for 25 or 30 min, and they found that consensus did not increase with acquaintance. However, researchers have found that consensus is higher among real-life acquaintances than among relative strangers—that is, people who viewed dyadic interactions for either 5 min (Funder & Colvin, 1988) or 25 or 30 min (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 among people who were unacquainted or had interacted only once (Kenny et al., 1994). In general, higher levels of consensus have been associated with higher acquaintance in studies with between-subjects designs in which there was a relatively large difference in the amount of available information between the high and low acquaintance groups, but not when there was a relatively small difference in acquaintance.

Between-subjects designs have generally found support for the acquaintanceship effect when self–other agreement is the outcome. For example, self–other agreement was higher among people who watched videotaped unstructured dyadic interactions for 25 to 30 min versus 5 or 10 min (Blackman & Funder, 1998), people who had been roommates for more than 10 months versus less than 10 months (Bernieri et al., 1994), people who indicated they knew the target of judgment extremely well versus not at all (Paunonen, 1989), and real-life acquaintances versus people who had watched 5-min unstructured dyadic interactions (Funder et al., 1995). In between-subjects designs, information quantity appears to have a positive linear relationship with self–other agreement across all levels of acquaintance. However, a complication common to between-subjects designs of consensus and self–other agreement is that real-life acquaintances have self-selected to know a person for longer, which introduces the possibility that confounds such as similarity to the target and liking of the target influence accuracy in addition to length of acquaintance. This is why, in the current study, we have experimentally manipulated information quantity.

In within-subjects research designs in which the same participants are involved at each level of acquaintance, previous findings

---

1 With contrast weights of –1, 0, +1 for RAM and –2, +1, +1 for WAM, the correlation between the predictions was .87.
suggest that as acquaintance increases, consensus generally remains constant, whereas 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-min group interactions on 4 consecutive days (Park & Judd, 1989), students who participated in weekly 20-min group meetings for 7 weeks (Paulhus & Bruce, 1992), and students living near each other in a dorm from 2 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 for between 8 min and 2 hr 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-min group meetings for 7 weeks (Paulhus & Bruce, 1992). Also, findings by Borkenau and colleagues (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004) suggest that acquaintance and accuracy are positively related when accuracy is operationalized as the level of agreement with ratings from the target, acquaintances of the target, experimenters, and a confederate who interacted with the target; and when acquaintance is operationalized as the number of behavioral episodes observed and rated by different judges that were included in a composite judgment.

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 (Kenny et al., 1994; Park & Judd, 1989; Park et al., 1997; Paulhus & Bruce, 1992) 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).

Another possible limitation of past research 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 concern was addressed by asking participants a series of factual questions about their interaction partners that sampled from the kinds of information people might learn in an initial interaction. This sample of relatively objective questions was used as a composite judgment that has been shown to capture the important aspects of personality information acquisition and realistic accuracy, consensus, and self–other agreement. The information quantity analyses are a conceptual replication of a study by Blackman and Funder (1998). The current project also went beyond Blackman and Funder’s study and other previous 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. This method may have different implications for the way judges detect and use information when making personality judgments. For example, judges watching videotapes have no need to be concerned with how they are perceived or what they will say next, whereas judges who are involved in an interaction are likely to be concerned with these and other issues. More important, many researchers, including Blackman and Funder, have used self–other agreement as a proxy for accuracy, whereas in the current study, we went beyond self–other agreement and examined 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 personality relevance of the information that becomes available within a given period of time. Intuitively, it seems apparent that not all kinds of information contribute equally to the achievement of an 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 was one goal of the current study.

Anderson’s (1984) groundbreaking study supported the idea that information quality is related to self–other agreement. Anderson found that observing an interview in which questions about thoughts and feelings were discussed yielded higher self–other agreement of personality ratings than did observing an interview in which hobbies and activities were discussed. One conclusion that can be based on this finding is that thoughts and feelings are higher quality information than hobbies and activities, and this is why the judges who learned about thoughts and feelings were able to agree more closely with the target’s self-judgments of personality than were the judges who learned about hobbies and activities.

Another aspect that might influence the quality of the information that is likely to be revealed 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 with 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.

The present study was designed to begin the investigation of what occurs when people interact in situations with different levels of information quality. A strong situation was created experimentally by giving participants the specific task of answering a long set of trivia questions. Two weaker situations were created by telling
participants to talk about whatever they would like or to get to know one another as well as possible. We predicted that more personality-relevant information would be revealed in the two weak situations than in the strong situation and that realistic accuracy, self–other agreement, and consensus would be higher in the two weak situations. We also predicted that more personality-relevant information would be available in the weak situation in which participants were told to get to know each other than in the weak situation in which participants were told to talk about whatever they would like, because the participants would be more likely to reveal or ask about personality-relevant information as they attempted to get to know each other than when they were simply trying to pass the time by talking about whatever they liked.

Research Questions

With the theoretical basis of both the RAM (Funder, 1995) and the WAM (Kenny, 1991) in mind, in the current study we examined 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 achieve higher levels of realistic accuracy, consensus, and self–other agreement than people who interact for shorter periods of time.

To examine this issue, we compared personality judgments following three-person unstructured interactions of three lengths (minimal interaction, 50 min, and 3 hr). On the basis of RAM (Funder, 1995, 1999), we expected that all three indicators of personality judgment—realistic accuracy, consensus, and self–other agreement—would increase linearly across these experimental conditions. On the basis of WAM (Kenny, 1991), we expected 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 were 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 than participants who interact in situations in which less personality-relevant information is likely to be available.

In one of the first experimental investigations to directly examine this issue, we compared the accuracy of personality judgments made following one of three 50-min interactions that vary in the amount of personality-relevant information likely to be available in them. The low-quality condition was a strong situation that allowed for little behavioral variation and was expected to elicit the least amount of personality-relevant information and therefore the lowest levels of realistic accuracy, consensus, and self–other agreement. The medium- and high-quality conditions were both relatively weak situations that were 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 were expected to share a broader range of personality-relevant information and to 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, on the basis of RAM, we predicted a linear relationship between information quality and all three indicators of accuracy. We could 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 acquaintance parameter of WAM should increase with information quality. Therefore, on the basis of WAM, we predicted that consensus would increase between the low- and medium-quality conditions and remain about the same between the medium- and high-quality conditions.

Method

The current study was a between-subjects experimental design in which each participant interacted in one of five conditions. The first independent variable, information quantity, had three levels: low (minimal information condition), medium (short unstructured condition), and high (long unstructured condition). The second independent variable, information quality, also had three levels: low (trivia condition), medium (short unstructured condition), and high (get to know condition). The effects of each independent variable were examined for three dependent variables: realistic accuracy (agreement between personality judgments and a broad-based accuracy criterion), self–other agreement (agreement between judgments and the target’s self-ratings), and consensus (agreement between two judges about a target).

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 men and 90 women) 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 was 38% Asian American, 20% Hispanic, 14% Caucasian, 12% African American, and 16% other or not specified. There were three groups of each possible gender composition (female–female–female [FFF], male–male–male [MMM], female–female–male [FFM], and female–male–male [FMM]) within each experimental condition, and all of the gender compositions were analyzed together because of the balanced nature of the design and because the number of participants within each gender composition would be too small to find reliable results if analyzed separately.

This is the second article 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, 2005), or with future planned projects.

Materials

California Adult Q (CAQ) set. The CAQ (Block, 1961, as modified for use by nonprofessionals by Bem & Funder, 1978) 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 (extremely uncharacteristic) to 9 (extremely characteristic). Participants rated themselves and their group interaction partners using the Q-item rating, and acquaintance informants also rated the target participants 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 judges 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 predetermined number of items.

Information fact sheet. An information fact sheet was constructed for RAP-II that consisted of open-ended questions that were designed to sample from the wide array of information that could become available during the experimental interactions. The information fact sheet was used as a manipulation check to determine whether participants in higher quantity and quality interactions actually acquired more information about their interaction partners than did participants in lower quantity or quality interactions. Questions included information about 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 correlate with 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 the ratings of personality and was completed on a later occasion by each participant about him- 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, as the decisions about the correctness of an answer were relatively objective, any discrepancies in scoring were reconciled by a third judge. The reliability of the 19 items was high enough to justify using the sum as a score representing information acquisition (Cronbach’s alpha = .77).

Procedures

Overview. 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 were 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 a subsequent session, each participant was interviewed by a clinical psychologist who described the personality of the participant after the conclusion of the interview. Additionally, acquaintance informants completed questionnaires regarding the target participant with whom they were acquainted.

Presession. At the presession, participants were given the first take-home packet of questionnaires and were scheduled into one of the five experimental conditions. The times of the conditions were established in advance, and the participants chose a session that fit with their schedules. Although technically this procedure was only quasi-random, it did assign participants to conditions independently of any aspect of the experimental design, and we have no reason to believe that participants in the various conditions differed from each other in any way that might have affected the outcome of the study. Additionally, 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. If the participant recognized another participant in the same group, the participant was assigned to a different group. This procedure ensured that none of the participants were acquainted with their interaction partners prior to the experimental group interaction.

Experimental conditions. Target participants interacted in the laboratory in one of five experimental conditions designed to vary in 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: all female (FFF), all male (MMM), 2 females and 1 male (FFM), and 1 female and 2 males (FM). An equal number of the four gender compositions was assigned to each experimental condition. A total sample size of 180 allowed for a balanced design with 3 groups of each gender composition per condition, for a total of 12 groups per condition.

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 three measures, which included the rating form of the CAQ and the information fact sheet.

Minimal information condition. Participants rated the other two participants in the room immediately following a short set of instructions explaining that they were studying first impressions and to do the best they could describing their 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 1 hr). This situation provided an empirical estimate of the baseline of realistic accuracy, consensus, and self- other agreement and was used as the low-level condition for information quantity.

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 min 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 would leave relatively little room for extraneous commentary or the disclosure of personality-relevant information, as the group would spend the entire interaction time responding to the trivia questions. This condition was used as the low-level condition for information quality.

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

Long unstructured condition. Participants were told that they could talk about anything they liked for the next 3 hr, 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 provided a weak situation with very little structure and should have included at least some personality-relevant information on which judgments could be based. This condition was used as the high-level condition for information quantity.

Get to know (GTK) condition. Participants were told that their task for the next 50 min was to get to know each other as well as possible and to learn as much as they could about what type of person each of them was.

2 On completion of data collection, it was found that 2 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. To maintain a fully balanced design, we report results from the data of all 60 groups.
This context provided a weak situation that was expected to yield a wide range of behaviors relevant to personality, and the objective of getting to know each other was expected to guide participants to probe for and to remember information related to personality. This condition was 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) interviewers who had experience with a college population. With consent of the participants, all interviews were videotaped. The clinicians conducted a 1-hr semistructured life history interview adapted from a protocol used for many years by the Institute of Personality Assessment and Research (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, “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 [their] life that had a significant impact on or changed [their] life in some way.” 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. When two ratings of the participant were available, a composite score was computed for each item and used in subsequent analyses. The average interclinician profile agreement was \( r = .50 \) (\( SD = 0.17 \)), and the coefficient of internal consistency based on a composite of the two raters was .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 was \( r = .40 \) (\( SD = 0.19 \)). The average coefficient of internal consistency, or the dependability of the profile based on a composite of the two acquaintances, was .57.

Analyses

Realistic accuracy criterion. Many researchers have used self–other agreement as a proxy for accuracy, but in the current project we strove 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 these ratings were used to compute a composite score for each item. First, when ratings from two acquaintances or two clinicians were available, these ratings were averaged. Then, a simple average for each item was computed across the three ratings from the self, acquaintances, and clinician. The construct of realistic accuracy is very new, so there is not an established level of reliability for determining the adequacy of the accuracy composite. However, on the basis of the magnitudes of correlations that are reported in the self–other agreement literature (Bernieri et al., 1994; Funder, 1980), the mean alpha reliability of the accuracy composite seemed high enough to justify the use of these average ratings (\( M = 0.42, SD = 0.13, \) range = 0.08–0.72).

Computation of realistic accuracy. Profile analysis allows for an examination of judgments in regard to a target’s overall personality by the use of the entire set of 100 CAQ items in a single analysis. The computation of a profile correlation simply involves correlating two sets of ratings for the same target across the 100 items of the CAQ. First, to examine realistic accuracy, we correlated the ratings by each judge of each interaction partner with the accuracy composite for that interaction partner, resulting in a profile correlation that represented the relation between the ratings of the judge and an approximation of what the target is really like. This procedure yielded six scores for each group (two scores for each of three judges). These scores could be transformed using Fisher’s \( r \)-to-\( z \) transformation and averaged so that each group had a single realistic accuracy score (see Figure 1).

Computation of consensus. To examine consensus, we correlated the ratings provided by the two interaction partners of each target with each other, resulting in a profile correlation that represented the degree of similarity in the partner’s ratings of that target. This analysis yielded three consensus scores for each group (one for each target), which could be \( Z \) transformed and averaged so that each group had one consensus score.

Computation of self–other agreement. Finally, to examine self–other agreement, we correlated the ratings of each interaction partner with the partner’s self-ratings, resulting in a profile correlation that represented the relation between the ratings of the judge and the target’s self-ratings. As in realistic accuracy, this procedure yielded six self–other agreement scores.

<table>
<thead>
<tr>
<th>CAQ 1</th>
<th>Rating of partner A</th>
<th>Accuracy Criterion</th>
<th>Rating of partner B</th>
<th>Accuracy Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>CAQ 100</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 1. Computation of profile realistic accuracy scores.](image)

3 The full protocol for the clinical interview is available on the Riverside Accuracy Project’s Web site at [http://www.rap.ucr.edu/interview.htm](http://www.rap.ucr.edu/interview.htm).

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

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

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

7 When ratings from only one acquaintance or clinician were available, then there was no need to compute an average before averaging the ratings with the other sources.

8 On the basis of correlations among types of raters, the accuracy criterion is appropriate to use for computing profile correlations: self–acquaintance (\( M = 0.48, SD = 0.17 \)), self–clinician (\( M = 0.39, SD = 0.14 \)), acquaintance–clinician (\( M = 0.37, SD = 0.19 \)), Spearman–Brown for three raters = .68.
per group, and these scores could be Z transformed and averaged so that
each group had a single self–other agreement score.

Components of accuracy scores. 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, which result from the use of response
sets or a reliance on stereotypes. The effects of response sets were not of
concern for between-groups comparisons in the current experiment be-
cause participants were randomly assigned to one of five experimental
conditions, and therefore, participants with various response sets were
equally distributed among the conditions. The effects of stereotype accu-
rate also were not of concern because stereotype accuracy should have
equivalent effects on the magnitude of the profile correlations across all
conditions, and we were interested in the differences between condi-
tions and not in the absolute magnitude of accuracy. For this reason, we used
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 was not expected to be zero, as
some items of the CAQ are generally rated higher than others, regardless of
the target (Blackman & Funder, 1998). To determine baseline profile
correlations within the current data, we computed the correlations among
30 random pairs of ratings. For realistic accuracy, the profile correlations
were computed between the ratings of one target and the accuracy com-
posite for a different target, which resulted in a baseline profile correlation
of .36. For consensus, the profile correlations were computed between two
interaction partners of different targets, which resulted in a baseline profile
correlation of .25. For self–other agreement, the profile correlations
3 were computed between the ratings of one target and the self-ratings of a
different target, which resulted in a baseline profile correlation of .28.
These baseline correlations may at first seem somewhat high, but when one
realizes that all participants in the sample were college students and that,
in general, college students are similar to each other in many ways, it
makes sense that even random pairings resulted in profile correlations of
moderate magnitude. Profile correlations greater than these baseline cor-
relations indicate differential accuracy (Blackman & Funder, 1998).

Contrast analysis. Contrast analysis was used to test focused questions
by assigning weights to experimental conditions that reflect a prediction
concerning the structure of the data (Rosenthal, Rosnow, & Rubin, 2000).
Once contrast weights were assigned to each condition, it was possible to
determine the extent to which the data were in line with the prediction.
When a t test was computed for a contrast analysis,9 the degrees of freedom
were the number of participants minus the number of groups.10 One-tailed
p values that are associated with the t contrast are reported. It was also
possible to determine how much variance in the group means was ac-
counted for by the prediction by correlating the group means with the
contrast weights and squaring the obtained correlation, which resulted in an
t2 alerting. A large t2 alerting provides support for the prediction. Furthermore,
it was possible to test whether the leftover variance, or the variance
not accounted for by the prediction, was significant by computing a
11 A nonsignificant t noncontrast also lends support to the a priori
prediction because little variance is left over that is not accounted for by the
prediction.

On the basis of RAM, we predicted that there would be linear relation-
ships among information quantity and quality and realistic accuracy, con-
sensus, and self–other agreement, for which the appropriate contrast
weights are −1 (low information), 0 (medium information), and +1 (high
information). On the basis of WAM, we predicted a slightly different
relationship between information quantity and quality and consensus,
which reflects an increase at low levels of acquaintance and then a leveling
off at higher levels of acquaintance, for which the appropriate contrast
weights are −2 (low information), +1 (medium information), and +1
(high information).

Post hoc t tests. Contrast analyses are informative regarding how well
the obtained data fit an a priori prediction about how the data should be
ordered, but they do not provide information regarding which pairs of
groups are reliably different from each other. Therefore, post hoc
independent-samples t tests were used to determine which pairs of exper-
imental conditions had reliably different levels of realistic accuracy, con-
sensus, and self–other agreement. No adjustments were made to the p
values associated with these tests, so the conservative reader may wish to
adjust the p values to account for the number of comparisons in each
analysis.

Results

The focus of our analyses was on realistic accuracy, as this
project was 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 also provided evidence concerning our predictions
for consensus and self–other agreement. Before examining the
effects of information quantity and quality on judgment accuracy,
we first determined whether participants in longer and higher
quality interactions did indeed acquire more information (on the
basis of a sample of information possibly available in such inter-
actions), as a manipulation check. Then, we used contrast analysis
(Rosenthal et al., 2000) to examine how well the changes in
accuracy across experimental conditions fit the predictions based
on RAM for realistic accuracy, consensus, and self–other agree-
ment and based on WAM for consensus only. Finally, the degree
to which information acquisition was related to realistic accuracy,
consensus, and self–other agreement, regardless of the type of
group the participants interacted in, was examined.

There are several possible ways to analyze these data. We used
traditional statistical procedures to examine differences at the
group level by conducting a between-groups analysis to determine
how well the obtained data fit the predictions. Our predictions
were concerned with properties of the context, not of the individ-
ual, and a group-level analysis reflects this orientation. Although
none of the basic conclusions of this study change when data are
examined at different levels of analysis, an advantage of analyzing
data at the level of the group is that it allowed us to perform all
analyses in a consistent manner (at the same level of analysis).

Information Acquisition

To determine whether participants in interactions of greater
quantity and quality acquired more information about each other,
we compared the scores achieved on the information fact sheet across experimental conditions. This can be considered a manipulation check that determines the degree to which participants in longer and higher quality interactions actually learned more about each other.

Information quantity. A good fit was found between amount of acquired information and a linear increase across the three levels of information quantity: minimal information ($M = 2.80, SD = 1.10$); short unstructured ($M = 5.17, SD = 1.09$); long unstructured ($M = 7.00, SD = 1.58$); and $r_{contrast} = .81, t_{contrast(33)} = 8.02, p < .0001$ (see Figure 2). Furthermore, on the basis of independent-samples $t$ tests, we found that the scores of all groups differed significantly from each other: minimal information and short unstructured, $t(22) = 5.29, p < .0001$; minimal information and long unstructured, $t(22) = 7.54, p < .0001$; and short unstructured and long unstructured, $t(22) = 3.29, p = .003$.

Information quality. A good fit was also found between the amount of acquired information and a linear increase across the three levels of information quality: trivia ($M = 1.65, SD = 1.04$); short unstructured ($M = 5.17, SD = 1.09$); GTK ($M = 5.87, SD = 1.57$); and $r_{contrast} = .82, t_{contrast(33)} = 8.22, p < .0001$ (see Figure 3). Participants in both the short unstructured and GTK conditions acquired reliably more information than did participants in the trivia condition, $t(22) = 8.08, p < .0001$, and $t(22) = 7.75, p < .0001$, respectively, although information acquisition did not differ reliably between the short unstructured and GTK conditions, $t(22) = 1.26, p = .22$.

These findings concerning a small sample of the available information imply that participants in longer and higher quality interactions acquired more information about each other. The next analyses addressed the question of whether personality judgments were also more accurate in the longer and higher quality interactions.

Personality Judgment

Information quantity. To test the prediction that judges would be more accurate after interacting for longer periods of time, we examined personality judgments at three levels of information quantity: low (minimal information—no interaction), medium (short unstructured—50-min interaction), and high (long unstructured—3-hr interaction). First, profile correlations were computed for each group, as described in the Analyses section. The profile correlations were transformed using Fisher’s $r$-to-$z$ transformation, and then contrast analyses (Rosenthal et al., 2000) were used to determine the degree to which the profile correlations followed the predictions based on RAM and WAM.

As predicted, we found a good fit between realistic accuracy scores and the prediction of a positive linear relationship with information quantity: minimal information (mean $r = .29, SD = 0.20$); short unstructured (mean $r = .45, SD = 0.14$); long unstructured (mean $r = .47, SD = 0.09$); and $r_{contrast} = .53, t_{contrast(33)} = 3.56, p = .000058$ (see Figure 4). The linear prediction accounted for 86% of the variance in the group means, and the variance not accounted for by the prediction did not reach the .05 level of significance, $t_{noncontrast(33)} = 1.41, p = .08$. When the groups were compared with each other, realistic accuracy was reliably higher in both the short unstructured and long unstructured conditions than in the minimal information condition, $t(22) = 2.57, p = .02$, and $t(22) = 3.46, p = .002$, respectively, but accuracy in the long unstructured condition was not reliably higher than accuracy in the short unstructured condition, $t(22) = 0.74, p = .47$.

Information quantity also had a good fit to the prediction of a positive linear relationship with consensus: minimal information (mean $r = .25, SD = 0.18$); short unstructured (mean $r = .41, SD = 0.20$); long unstructured (mean $r = .44, SD = 0.17$); and $r_{contrast} = .44, t_{contrast(33)} = 2.83, p = .004$. The linear prediction accounted for 86% of the variance in the group means, and the
variance not accounted for by the prediction was not significant, $t_{\text{noncontrast}}(33) = 1.14, p = .13$. Judges in the short unstructured and long unstructured conditions had reliably higher consensus than judges in the minimal information condition, $t(22) = 2.33, p = .03$, and $t(22) = 2.98, p = .007$, respectively, although consensus did not differ for judges in the short unstructured and long unstructured conditions, $t(22) = 0.42, p = .68$.

The consensus data were also in line with the prediction based on WAM, in which an increase was predicted between the low- and medium-quantity conditions but not between the medium- and high-quality conditions, $r_{\text{contrast}} = .48, t_{\text{contrast}}(33) = 3.04, p = .0023$. The nonlinear prediction accounted for 98% of the variance in the group means, and the variance not accounted for by the prediction was not significant, $t_{\text{noncontrast}}(33) = 1.18, p = .12$.

Finally, information quantity had a good fit to the prediction of a positive linear increase with self–other agreement: minimal information (mean $r = .21, SD = .19$); short unstructured (mean $r = .36, SD = .12$); long unstructured (mean $r = .39, SD = .12$); $r_{\text{contrast}} = .49, t_{\text{contrast}}(33) = 3.24, p = .0014$. The linear prediction accounted for 86% of the variance in the group means, and the variance not accounted for by the prediction was not significant, $t_{\text{noncontrast}}(33) = 1.29, p = .10$. Judges in the short unstructured and long unstructured conditions had reliably higher self–other agreement than judges in the minimal information condition, $t(22) = 2.58, p = .02$, and $t(22) = 3.03, p = .006$, respectively, although self–other agreement did not differ for judges in the short unstructured and long unstructured conditions, $t(22) = 0.72, p = .55$.

Note that for all indicators of accuracy, the mean profile correlations for the minimal information condition were approximately equal to or smaller than the random baseline correlations of .36 for realistic accuracy, $t(11) = 0.01, p = .99$; .25 for consensus, $t(11) = 0.11, p = .91$; and .28 for self–other agreement, $t(11) = 1.29, p = .22$; but the mean profile correlations were all higher than the random baseline correlations for the short unstructured condition: realistic accuracy, $t(11) = 2.94, p = .01$; consensus, $t(11) = 3.22, p = .008$; and self–other agreement, $t(11) = 2.76, p = .02$; and for the long unstructured condition: realistic accuracy, $t(11) = 6.31, p < .0001$; consensus, $t(11) = 4.44, p = .001$; and self–other agreement, $t(11) = 3.53, p = .005$. 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 the relevance of information to personality. To test our prediction that judgments would be more accurate when participants interacted in higher quality situations, we had participants interact 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 to test the prediction based on RAM and a nonlinear relationship to test the prediction for consensus based on WAM.

As predicted, we found that realistic accuracy had a good fit to the prediction of a positive linear relationship with information quality: trivia (mean $r = .36, SD = .12$); short unstructured (mean $r = .45, SD = .12$); GTK (mean $r = .49, SD = .16$); and $r_{\text{contrast}} = .43, t_{\text{contrast}}(33) = 2.71, p = .0053$ (see Figure 5). The linear prediction accounted for 97% of the variance in the group means, and the variance not accounted for by the prediction was not significant, $t_{\text{noncontrast}}(33) = 0.50, p = .31$. Realistic accuracy was reliably higher in the GTK condition than in the trivia condition, $t(22) = 2.76, p = .01$, and the difference between the trivia and short unstructured conditions approached the conventional level of significance, $t(22) = 1.94, p = .07$. However, the GTK condition did not differ reliably from the short unstructured condition, $t(22) = 0.88, p = .39$.

A good fit to the prediction of a linear increase was also found with consensus: trivia (mean $r = .29, SD = .17$); short unstructured (mean $r = .41, SD = .20$); GTK (mean $r = .41, SD = .10$); and $r_{\text{contrast}} = .33, t_{\text{contrast}}(33) = 1.99, p = .028$. The linear prediction accounted for 74% of the variance in the group means, and the variance not accounted for by the prediction was not significant, $t_{\text{noncontrast}}(33) = 1.19, p = .12$. Judges in the GTK condition had reliably higher consensus than judges in the trivia condition, $t(22) = 2.34, p = .03$; the difference between consensus in the short unstructured and trivia conditions approached significance, $t(22) = 1.78, p = .09$; and consensus did not differ for judges in the short unstructured and GTK conditions, $t(22) = 0.04, p = .97$.

The consensus data were also in line with the prediction based on WAM, $r_{\text{contrast}} = .37, t_{\text{contrast}}(33) = 2.31, p = .014$. The nonlinear prediction accounted for 99% of the variance in the group means, and the variance not accounted for by the prediction was not significant, $t_{\text{noncontrast}}(33) = 0.001, p = .50$.

Finally, a good fit to the prediction of a linear increase was found with self–other agreement: trivia (mean $r = .27, SD = .09$); short unstructured (mean $r = .36, SD = .12$); GTK (mean $r = .38, SD = .16$); $r_{\text{contrast}} = .39, t_{\text{contrast}}(33) = 2.40, p = .011$. The linear prediction accounted for 91% of the variance in the group means, and the variance not accounted for by the prediction was not significant, $t_{\text{noncontrast}} = 0.75, p = .23$. Judges in the short unstructured and GTK conditions had reliably higher self–other agreement than judges in the trivia condition, $t(22) = 2.16, p = .04$, and $t(22) = 2.33, p = .03$, respectively, although self–other agreement did not differ for judges in the short unstructured and GTK conditions, $t(22) = 0.49, p = .63$.

Note that again, the mean profile correlations for the trivia condition were approximately equal to or smaller than the random baseline correlations of .36 for realistic accuracy, $t(11) = 0.48$,

![Figure 5. Mean realistic accuracy scores for experimental conditions used in the information quality analyses. Note that the random baseline profile correlation was .36. GTK = get to know condition.](image-url)
Relation Between Information Acquisition and Accuracy

We could 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, regardless of the type of experimental condition in which the participants interacted.12 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 five 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 regardless of information quantity or quality, groups that acquired more of the objective information that was sampled with the information fact sheet also judged their partners with higher average levels of accuracy, consensus, and self–other agreement.

Discussion

The preceding analyses provide tests of the relationships between two aspects of good information and the accuracy of personality judgment. Experimentally manipulated information quantity, or length of acquaintance, was positively related to realistic accuracy, consensus, and self–other agreement. Additionally, 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 Quality

RAM predicts that accuracy will increase in relation to information quantity (Funder, 1995), but previous research has only tested this prediction using self–other agreement, which is a limited definition of accuracy. The current research confirms past findings that self–other agreement increases with acquaintance (Bernieri et al., 1994; Blackman & Funder, 1998; Funder et al., 1995; Paulhus & Bruce, 1992; Paunonen, 1989) and provides one more piece of evidence in favor of the existence of a 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 our prediction was that consensus would increase at low levels of acquaintance and then stay about the same even as acquaintance continued to increase (Kenny, 1991). These predictions are fairly similar, and we found support for both predictions. It is not possible to determine if one prediction is superior to the other, for a couple of reasons. First, the predictions themselves are highly similar, and so any data should fit both predictions to about the same degree. Second, it was not possible to determine whether the experimentally manipulated difference between the low- and medium-information quantity conditions was equal to the difference between the medium and high conditions and, therefore, whether a linear or nonlinear prediction was more appropriate. However, it is obvious from the means of the conditions that the larger increase was between the low and medium groups, but it is not possible to determine whether the findings support the prediction based on WAM better than the prediction based on RAM.

Overall, these findings support the existence of the acquaintanceship effect and the role of information quantity in accurate judgment. This interpretation was bolstered by the additional finding that participants learned more facts about each other in the longer interactions and therefore had a larger quantity of information available to them when making judgments of personality.

Information Quality

To our knowledge, the current article 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 and instructions that provided 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, whereas 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 should have increased 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, 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 was available to be detected (encoding), that the detected information was remembered and utilized better (decoding), or both.

As noted in the Results section, even though the data showed a good fit to a linear prediction, most of the increase in the indicators of accuracy was between the low- and medium-quality conditions, and the medium- and high-quality conditions were quite similar. One explanation for this finding is that people naturally try to get to know each other in an initial interaction, so even though we explicitly instructed participants in the high-quality condition to get to know each other and simply told participants in the medium-quality condition to talk about whatever they liked, they may have had similar interactions in which they learned approximately the same amounts of personality-relevant information that could be used when making judgments of personality. If this is the case, then information quality may not have been very different between our medium- and high-quality conditions, which would explain why realistic accuracy, consensus, and self–other agreement were also not very different in these conditions.

Limitations and Future Directions

As in any study, our experimental design has limitations that may affect interpretation. First, the construct of realistic accuracy is a philosophical ideal, and any measurement of it must necessarily be an approximation. We attempted to approach realistic accuracy by having several knowledgeable people, including expert judges, describe the personalities of the target participants. This procedure was not likely to tell us exactly what the target is like, but it was 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 interpreted as evidence that realistic accuracy cannot be meaningfully distinguished from the simpler and more traditional operationalizations or that constructing multifaceted criteria is not worth the time and effort. We 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 in which 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 such differentiation between constructs 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 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 information or information that is highly relevant to personality.

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 1st-year freshmen in college dorms, participants in volunteer corps who live and/or work together (i.e., Peace Corps, Jesuit Volunteer Corps), and older adults 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.
Conclusion

The current findings support several aspects of RAM, including the implications of information for the relevance, availability, and detection stages that make accurate judgment possible, along with specific predictions of WAM concerning consensus. Our main findings are that 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 for or encourage people to reveal personality-relevant information.

References


