Information as a Moderator of Accuracy in Personality Judgment

Andrew Beer

University of South Carolina Upstate

Abstract

People inherently believe that additional information is helpful in making accurate personality judgment, an assertion supported by empirical evidence. In this chapter, I review the evidence beginning with the cross-sectional and longitudinal study of accuracy in naturally-existing groups and continuing through to laboratory-based experiments involving the intentional manipulation of available information. In doing so, I discuss the process of becoming acquainted with others in our social world and make suggestions for future avenues of research in this area, including but not limited to more clearly defining acquaintanceship, studying information quantity and quality jointly and separately, and better connecting personality judgment with real-world phenomena.

Key Words: accuracy, information quantity, information quality, acquaintanceship, personality judgment

Information as a Moderator of Accuracy in Personality Judgment

In most modern cultures, it would be considered reckless to propose marriage after a few weeks of dating and brazen to ask for a promotion on one's third day on the job. However, after a considerable amount of time and exposure to the relevant parties, such requests are commonplace and expected. It is natural to want to truly know a person before tying one's lot to theirs, personally or professionally, and one could argue that in many ways, knowing a person means that we feel able to accurately judge his or her personality. Evidence of the belief that greater information tends to associate with greater accuracy in personality judgment is easily available. Employers seek letters of reference from individuals who *should* know the applicant, usually based on length and depth of exposure—often specifying the length and nature of such exposure. In social settings, we make general assumptions about our knowledge of others, and this is often based on our sense of appropriate exposure (e.g., "I only know Ted from work; you might ask his brother if he would like tickets to the opera.").

In an effort to understand the psychological underpinnings of accurate personality judgment, Funder's Realistic Accuracy Model (RAM; 1995; 2012; see also Chapter 2 in this handbook) envisions the path to accurate judgment as occurring through four successive processes. First, potential cues must be *relevant* to an underlying personality characteristic of interest. Second, these relevant cues must be *available* to the perceiver in question. Third, the perceiver must *detect* these cues, and finally, the perceiver must properly connect these cues to the relevant personality characteristic, a process called *utilization*. One of the principal moderators of accuracy in RAM is Information, which works primarily through the first two processes of RAM. That is, *good information* means that the judgment context is characterized by greater availability of relevant cues to personality. More and better information helps us make more accurate personality judgments, but what do "more" and "better" mean in these circumstances? What, exactly, happens during the course of dating or during an occupational

performance evaluation period? Do we come to know others via repeated bouts of small talk, or does most useful information come from long, focused conversations about important topics? Can we come to understand someone simply by sharing a cubicle wall for two years, without forming a meaningful relationship? How would this occur? In this chapter, I will trace empirical investigation of these questions from their origins to our present understanding of the phenomena, spotlighting a few landmark studies and concluding with some open questions and suggested directions for future research.

Origins

Research focused on the validity of observer judgments of personality has existed for the bulk of psychological scientific history (e.g., Cleeton & Knight, 1924; Estes, 1938; Shen, 1925). Almost as soon as the field began addressing research questions with data in earnest, attention turned towards identifying conditions that might yield accurate judgments of others—and determining if this were even truly possible. One of the earlier and most-cited empirical investigations into information's connection to accurate personality judgment, however, came rather incidentally. Norman and Goldberg (1966) were attempting to address a different issue (identifying principal personality traits) when they compared the inter-relations of trait judgments derived from statistical simulations to those obtained from groups of people making judgments of others. These sets of inter-trait correlations were similar, but it was their secondary tests that are relevant to the topic in question. In order to determine whether relations among trait judgments were due to actual differences in a target's personality (as opposed to the simple semantic relations of trait terms in the mind of the perceiver), Norman and Goldberg (1966) examined the two primary types of interjudge agreement: consensus and self-other agreement. Consensus involves the comparison of judgments made by two or more individuals (not including the person being judged, who is referred to as the target). Self-other agreement involves comparing a target's self-rated personality to judgments made by others. The simulated data were compared to data obtained from

four distinct groups: 1) strangers (participants made personality ratings of each other on the first day of class), 2) ROTC members (these people had typically seen each other in class or drills over a 1-2 year period), 3) Peace Corps trainees (these individuals had spent 3 months of intensive training together), and 4) fraternity seniors (this group had generally lived together over a period of 1-3 years). In this case, simulated data yielded no consensus (which was unsurprising, as the simulated "perceivers'" ratings were initially determined by a random response from which an algorithm generated related trait judgments), but groups with varying levels of acquaintanceship produced different levels of consensus. Specifically, the greatest levels of agreement were found in the fraternity and Peace Corps samples, followed by the ROTC sample and, finally, the strangers who showed the lowest levels of consensus among real (as opposed to simulated) data. The authors also presented self-other agreement data for the strangers and the Peace Corps sample, demonstrating that the Peace Corps trainees were more accurate in assessing each other's personalities (mean *r* across FFM traits = .41) than were the strangers (mean *r* = .25). At the time, the authors (appropriately) concluded that this was solid evidence in support of the existence of personality traits in the structure suggested, but—as they would note in the discussion--some of the additional findings of this work would spur entire areas of discovery in personality perception.

The Acquaintanceship Effect

The groups in Norman and Goldberg's (1966) study could be arranged in terms of increasing levels of subjectively determined "acquaintanceship". People who had lived together for years probably knew each other better than did people who spent a couple of hours a week together over a similar period of time, and each of these groups certainly knew more about each other than did people who had just met moments prior to evaluation. But these differences were not purely tied to mere length of exposure. For example, the manner in which Peace Corps trainees and ROTC members interacted was likely qualitatively different as well. Nevertheless, on its face the acquaintanceship effect seems rather

unexciting: those with greater personal knowledge of another should better understand that individual's personality, fostering more accurate personality judgments.

Much of the early empirical work directly addressing the phenomenon was cross-sectional in nature. At the same time that Norman and Goldberg reported their results, Taft (1966)—following pioneering work by Ferguson (1949) that demonstrated greater consensus in personality judgments made by employees of managers as they were better acquainted—undertook a more intentional look at the phenomenon. In Taft's study, participants nominated most- and least-known classmates, and self-other agreement was compared across these groups. As was the case in Norman and Goldberg's data, people were more accurate in judging the most known (r = .52) than the least known (r = .42) participants. Cloyd (1977) replicated these findings using a similar method (but different analytic strategy). In each of these studies, however, the samples were fairly small and the nomination procedures were somewhat problematic. In asking people to assess those they know, it is possible that these respondents were actually identifying people whom they thought they could best evaluate, in which case the exercise becomes somewhat redundant or at least an exercise in demonstrating a connection between perceived and actual accuracy rather than demonstrating the pure influence of level of acquaintance. A similar criticism could be made for any study in which acquaintanceship is operationalized via self-reported knowledge of the target, as was the case in both Watson and Clark's (1991) and Paunonen's (1989) work demonstrating support for the acquaintanceship effect. Even so, some studies utilizing such methodology showed only modest support for increased accuracy with increased acquaintance (Biesanz, West, & Millevoi, 2007).

In other work, researchers sidestepped this issue by employing methods more similar to Norman and Goldberg's (1966), in which consensus or agreement was compared across existing groups who differed in nature of relationship (and thus acquaintanceship). Funder and Colvin (1988) employed a unique design in which the same target was rated by each class of acquaintanceship,

diminishing some concerns common in such cross-sectional comparisons—and found that, indeed, friends (mean profile *r* = .46) were more accurate than strangers (*r* = -.03). Later research would replicate this effect using a similar design and variable-centered[1] analyses (Funder, Kolar, & Blackman, 1995). Getting beyond simple acquainted-unacquainted differences, Watson, Hubbard, and Wiese (2000) compared self-other agreement across groups of pairs of friendship, dating couples, and married couples. In this case, the first two groups had known each other for approximately 3 years versus an average of 17 years of marriage. As hypothesized, agreement was stronger among the married couples (mean *r* = .43) than among friends (mean *r* = .33) or dating couples (mean *r* = .35). Allik, De Vries, and Realo (2016) also demonstrated increased self-other agreement across groups characterized by increasing levels of familiarity or intimacy (friends, family, spouses or partners).

In the end, however, all of these data are subject to a myriad of confounds with respect to explaining differences in accuracy in terms of differences in information. In studies comparing groups, the issues are particularly stark. Beyond the quantitative (length of relationship) differences, are people who stay married for 10 years qualitatively different than dating couples? Married samples tend to be older, and it may be possible that older (or, more socially experienced) individuals are simply better judges of others. Or perhaps staying married for many years requires that one is a good judge of personality (Rogers & Biesanz, 2018; see also Ch. 6 by Colman in this handbook) ---or a good target (Human & Biesanz, 2013; see also Chapter 7 by Mignault & Human in this handbook). Further, intact

[1] Person-centered analyses involve comparing patterns of values across items (or traits) within a dyad, resulting in what are commonly referred to as profile correlations. Thus, when using a person-centered analytic strategy, one may calculate an accuracy score that can be applied to a given case or person but that is not as easily summarized within a trait category. Variable-centered analyses compare patterns of values across an entire sample for a given item (or trait). Thus, one calculates the accuracy in a given sample for Extraversion, for example. Biesanz, West, and Millevoi (2007) suggest in their discussion that perhaps different analytic strategies may yield different results with respect to the acquaintanceship effect, and while they are generally correct that different analytic strategies and foci yield different results--particularly when using person-centered analyses (e.g., Human & Biesanz, 2012) and distinguishing between normative and distinctive accuracy (e.g., Biesanz & Human, 2010)—the work reviewed here points to similar general conclusions regarding the effects of information quantity and quality on judgmental accuracy when using either person- or variable-centered analytic strategies.

groups or dyads of any sort may be more likely to be comprised of well-adjusted individuals, and psychological adjustment is a predictor of both good judges (Letzring, 2008, 2015) and good targets (Human & Biesanz, 2011). Additionally, in studies involving existing relationships it is possible that people who are acquainted have communicated with each other specifically about their personalities, thus agreement may reflect a simple result of that communication rather than truly accurate personality judgment (but see Funder, Kolar, & Blackman, 1995). These are just a few examples highlighting the difficulty in interpreting the results discussed previously. In order to alleviate some of these concerns—particularly those involving group differences—longitudinal studies are necessary.

Longitudinal Studies of Acquaintanceship

Most research on judgmental accuracy in relationships that develop over time takes a fairly standard form. As an initial example, consider a frequently cited study (Paulhus & Bruce, 1992) tracking the validity of observer judgments among previously unacquainted individuals. In short, students in a personality course met weekly in small groups 7 times for approximately 20 minutes per meeting. After each meeting, participants rated the personalities of the other members of their group. In keeping with expectations, self-other agreement showed a positive trend across the seven meetings: average (across FFM traits) correlations increased from r = .21 to r = .30 from the first to the last week. Kurtz and Sherker (2003) replicated this general effect with a sample of college roommates whom they assessed at approximately 2 weeks (mean r = .27) and 15 weeks (mean r = .43) of acquaintanceship. In the most recent attempt to examine the acquaintanceship effect over time, Brown and Bernieri (2017) assessed personality at zero acquaintance, after five minutes of interaction, and again after 10 weeks (a total of approximately 30 hours) of contact designed by the experimenters to create opportunities to become better acquainted (e.g., playing games, traveling short distances together, debating an issue, cleaning one another's rooms). They found general gains in accuracy from zero acquaintance to five minutes and from five minutes to ten weeks, although the gains in accuracy across the two time intervals were more

similar in magnitude than one might expect: the first five minutes seemed to provide as much relevant information as the next thirty hours! On the other hand, Park and Ryan's (1997) similarly designed—but longer (approximately 8 months at final assessment)—study revealed no increases in consensus with increased acquaintanceship. In fact, several studies of this sort (more on this later) yielded little in terms of support for increased information being associated with increased levels of consensus in personality judgment.

So why have longitudinal studies generated such a mixed bag of findings? One issue could be that very few studies start at a true zero point in terms of acquaintanceship. That is, the individuals in these groups have generally already spent some time together prior to the initial judgment of each other. As discussed later, this may be problematic for attempts to observe increasing accuracy, due to the fact that in some cases, very limited information can lead to accurate personality judgment. A second issue could be that these groups have tended to be student populations who are either thrown together by chance into work or discussion groups or even as roommates, which may not be fully representative of the natural flow of information and acquaintanceship in that it may lack the inherent motivational component of naturally forming relationships (i.e., the interactants truly care and intend to know each other). A third issue is that there tends to be fairly little control of exactly how information is being exchanged in these studies in terms of the kind of contact that is occurring and whether people are actually being exposed to more personality-relevant information.

The (Relatively Fruitless) Search for Acquaintanceship Moderators

In both cross-sectional and longitudinal studies of acquaintanceship effects, researchers have occasionally attempted to measure the precise methods by which people become acquainted. As mentioned previously, some early studies used self-reported "extent of knowledge" measures to create their comparison groups (e.g., Watson & Clark, 1991). In these studies, there is usually a relation

between knowing and accuracy, but as discussed previously, it is possible that the extent-of-knowledge measures are driven by an existing sense in the perceiver of self-other agreement with the target in question. Thus, others have chosen to measure acquaintanceship by more objective means. One way to do this is simply asking judges to estimate the length of their relationship with a given target. This has been operationalized in various ways, but generally has not yielded strong associations between length of acquaintance and degree of accuracy (e.g., Allik, De Vries, & Realo, 2016; Biesanz, West, & Millevoi, 2007). Indeed, there are reasons to believe that just the duration of a relationship may not, in fact, be a valid measure of acquaintanceship as it is intended in this line of research. Thus, others have attempted to more clearly measure acquaintanceship. In addition to measuring length of acquaintanceship, Watson, Hubbard, and Wiese (2000) used a modification of the Relationship Closeness Inventory (Berscheid, Snyder, & Omoto, 1989) and an ad-hoc measure in which they calculated the number of shared activities (e.g., going to movies, visiting family, exercising), in their "disappointing" (p. 555) attempts to uncover a statistical moderator of accuracy in their sample. Such unfulfilled hopes are not uncommon (e.g., McCrae, 1994; McCrae, Stone, Fagan, & Costa, 1998). In their discussion of moderators of accuracy in judgments of life satisfaction, Schneider, Schimmack, Petrican, and Walker (2010) noted other issues that may have hindered identification of acquaintance duration as a moderator of accuracy—most notably that (a) effect sizes of the relationship between accuracy and length of acquaintance are small (and thus require large sample sizes) and (b) common statistical techniques for identifying these relationships assume a linear connection between variables, which, in their estimation, is unlikely in the case of acquaintance effects (i.e., the learning curve is steeper sooner).

In short, attempts to establish acquaintanceship as a moderator of accurate personality judgment have been moderately successful when examining the phenomenon across groups but less successful when examining the phenomenon over time. Further, attempts to specifically measure the

way in which groups of varying levels of assumed acquaintanceship differ, and connect those differences to differences in accuracy, have also been generally unsuccessful. The various attempts and discussions thereof seem to imply that such failures are primarily measurement-based—if only we had clear, valid measures of acquaintanceship, the connection between it and accuracy would be empirically established to a greater extent. Therein lies the problem, however: prior to the development of a valid operational definition of acquaintanceship, there must be a consensual, coherent, conceptual definition of the concept. Perhaps we do not have a great sense of what is meant by "acquaintanceship" in its various applications.

A Model of Acquaintanceship

We have discussed the issues with simply relying upon direct measures of acquaintanceship (e.g., How well do you know this person?), but what is to take its place? How do the members of the Peace Corps and fraternity seniors differ? What about married couples and friends? In graduate school, I had one friend with whom I routinely played basketball, another with whom I would cook and play tennis, and another with whom I would discuss and exchange music. With whom was I most acquainted? For the first few years that I worked at my current job, a geology professor from down the hall would come by once a week to tell me a joke. Around that same time in my life, I spent the better part of a day talking with someone at a conference; I have not spoken to or seen that person since. With whom am I more acquainted? What really differentiates these relationships? What is the best way to conceptualize and measure acquaintanceship? Luckily, some researchers have attempted to address these questions empirically. Starzyk, Holden, Fabrigar, and McDonald (2006) asked samples of individuals to characterize their relationships with others of varying degree of subjectively-assessed "acquaintanceship"[2]. To uncover dimensions of acquaintanceship, they factor analyzed responses to a

[2] In this case, there were three levels: low (brief duration or homogenous interactions), moderate (moderate duration and/or varied interactions), high (long duration and varied types of interaction).

lengthy list of items ranging from queries about amount of contact ("I see this person a lot") to context-specific knowledge ("I know how this person handles stress") to simple behaviors ("I hug this person a lot"). From a list of over two hundred items, the authors identified six underlying dimensions. The first two seemed to tap quantity of information and reflect some simple conceptualizations of acquaintanceship: Duration (how long one has known the person) and Frequency (how often one sees the person). Duration is probably the more common measure utilized in the literature to date as a proxy for acquaintanceship, but it is important to distinguish it from frequency. For example, I have a friend whom I have known for over thirty years, but whom I see or speak to fairly rarely these days (long duration/low frequency), whereas I see my colleagues at work daily but have known them for a (relatively) short period of time (high frequency/short duration). Norman and Goldberg (1966), and many others that followed, were obviously sensitive to this difference in their chosen comparison groups, but they were also implicitly mindful of other distinctions. Fraternity seniors had lived together; Peace Corps volunteers likely had similar values (and lived together in that instance). Friends share information with each other and see each other in a variety of contexts, but dating and married couples are physically intimate and may have even more access to private information. These kinds of differences--which might be considered more qualitative in nature--are captured in the four other dimensions identified in Starzyk et al.'s research: Knowledge of Goals (extent to which one feels familiar with another's goals and interests), Physical Intimacy (extent to which physical contact is affectionate and common in the relationship), Self-Disclosure (familiarity with the other person's true feelings), and Social Network Familiarity (knowledge of the other person's friends and the interactions among them). Thus, one can also characterize level of acquaintance by evaluating a given relationship across these dimensions. For example, my relationship with the jokester was of moderate duration, low-to-moderate frequency, low self-disclosure, low physical intimacy, low knowledge of goals, and low-to-moderate knowledge of social network. The total relationship score across the six dimensions predicted

(at least modestly) self-other agreement in the original study, but has yet to be widely applied in research focused on accuracy in personality judgment (for various applications see Fareri, Niznikiewicz, Lee, & Delgado, 2012; Gros, Simms, & Antony, 2010; Nauta, 2012; Kahn, Hucke, Bradley, Glinski, & Malak, 2012; Sparling & Cramer, 2015;). Such empirically-derived measures of acquaintanceship might be of substantial utility in future work in this area.

The Last Word on Acquaintanceship (Meta-Analyses)

The studies summarized thus far were designed to evaluate differences in accuracy across naturally occurring groups of varying levels of acquaintance. Thus, each study had an aim to draw conclusions about how more information was associated with more accurate personality judgment. And despite the longstanding interest in conditions giving rise to accurate personality judgment, the studies aimed directly at answering this question could be considered relatively few. Happily, however, there is a wealth of accuracy data that can be compared in this fashion via meta-analytic review—even if it was not the authors' original intent to examine this particular question. As researchers in our field and others have begun to recognize the limitations of any single study (Maxwell, Lau, & Howard, 2015), the importance of such studies has appropriately grown.

An early meta-analytic review (Kenny, Albright, Kashy, & Malloy, 1994) indicated that increasing amounts of contact did not predict greater consensus among judges of personality in longitudinal designs (they did find evidence for increasing consensus with increasing acquaintance in cross-sectional designs). However, it must be noted that this review was (a) limited to studies utilizing specific design parameters, such as each judge rating multiple targets and each target being rated by multiple judges (limiting the number of studies included to 32) and (b) focused solely on consensus. The latter can be considered problematic in that two or more judges can agree on the nature of a given target while being entirely inaccurate. For example, two people could share the stereotype that males are more

aggressive, and thus agree that Ted is more aggressive than Suzy, even if this is not true. Thus, increased exposure may not always lead to gains in consensus, though it should generally lead to gains in accuracy (Blackman & Funder, 1998). Indeed, some models (e.g., the Weighted Average Model (WAM); Kenny, 1991; see also Chapter 3 by Malloy in this handbook) would not predict consensus to increase under the conditions observed in that analysis. Kenny and West's (2010) meta-analysis alleviates the concern that results may differ when using different accuracy measures by examining inter-judge agreement (a) involving the self and (b) not involving the self. Their conclusions with respect to information's moderating role on accuracy are the same, however: increased familiarity with a target does not seem to predict greater consensus or self-other agreement. It is again worth noting that this review was limited to round-robin designs, which allowed for the application of certain componential analyses (e.g., WAM; Kenny, 1991; again see Chapter 3 of this volume), and thus was limited to 24 studies.

In their meta-analysis on the utility of informant ratings of personality in general, Connelly and Ones (2010) used a more inclusive study selection strategy. They examined consensus, self-other agreement, and some instances of behavioral prediction across over 250 studies (comprising over 44,000 participants). Additionally, the authors used Starzyk et al.'s (2006) model of acquaintanceship as a means of organizing their analyses. Figure 1 presents the average self-other agreement correlations[3] across traits and across groups representing varying levels of acquaintanceship. As evidenced in the figure, the authors conclude that increased acquaintanceship is associated with increased accuracy (in the form of self-other agreement). Given its greater inclusivity, this study's result may be the more appropriate synopsis of the relationship between acquaintanceship and accuracy. That said, there is

[3] These are averaged across traits. To do so, I simply used Fisher's *z*-transformation on the average raw correlations presented in Table 5 of the original paper. The authors focus their conclusions on a different estimate, corrected for unreliability in both judgments, but I chose to use the raw correlations in keeping with other estimates provided in the text.

much to be considered regarding the mechanism of acquaintance. In Kenny et al.'s (1994) review, it was noteworthy that longitudinal studies, in particular, showed very little evidence of greater information leading to greater consensus. Indeed, Connelly and Ones (2010) suggest that sheer information quantity (e.g., greater exposure to target behavior) is only useful to a point, after which they posit that the self-disclosure associated with closer relationships is what drives greater accuracy in personality judgment. In other words, the *quality* of the information explains differences in accuracy between say, a spouse and a long-term co-worker. The data discussed to this point cannot directly address the mechanism of acquaintance, which is at the heart of inquiry with respect to information as a moderator of personality accuracy. In the remainder of this chapter, we will turn our focus toward specific attempts to more carefully control aspects of judgment context, and the distinction between quantity and quality of information will figure prominently in these discussions.

Beyond Acquaintanceship: Out of the Field and into the Laboratory

Much—though not all—of the work discussed to this point has involved examining differences in accuracy either (a) across naturally occurring groups of varying acquaintance or (b) within groups that are followed over a period of time with limited control of how information is exchanged. This lack of control makes inferences with respect to mechanism problematic. For example, in Paulhus and Bruce's (1992) discussion groups, one cannot be sure the extent to which self-disclosure occurred from week to week or what kinds of behaviors occurred in these contexts. The only sure thing is that *more* information has been exchanged. In order to better understand exactly how gains in accuracy might occur, it is necessary to exert more control over the availability of information in studies of judgment accuracy. In this section, I will review some of the systematic attempts to evaluate the role of information quantity and quality in the early stages of impression formation.

Zero Acquaintance: In the Beginning, There is Nothing (sort of)

Researchers have adopted the term "zero acquaintance" to describe studies in which judgments are made about individuals with whom we are unfamiliar—usually restricted to having met for the first time just prior to assessing personality. All relationships begin at a baseline of no knowledge—at least in theory. In practice, this is fairly rare, however. By the time we have met someone in person, we generally have a fair amount of potentially relevant information about them. For example, I "knew" my wife's best friend for years before having met her, so ground zero of person knowledge in this case would have to pre-date anything generally referred to as "acquaintance". Zero acquaintance and zero information are not necessarily the same thing. Early research (e.g., Norman & Goldberg, 1966) would (inadvertently) highlight this point in that unacquainted individuals can nonetheless form somewhat accurate impressions of one another.

One thing that might have struck the reader in preceding sections was the magnitude of accuracy estimates for strangers across the various studies. For traits like Extraversion, self-other agreement correlations in samples of strangers routinely reach r = .30 or greater. This level of agreement among strangers has also struck some researchers as a bit odd. In fact, though Norman and Goldberg (1966) did not spend time discussing this peculiarity in their data, Watson (1989) found it so unexpected that he conducted a study intended primarily to replicate these surprising zero acquaintance effects (Norman and Goldberg's strangers also generated statistically significant agreement correlations for Agreeableness and Openness/Culture). Watson (1989) indeed replicated the effect, as have others over time (e.g., Beer & Watson, 2008; Beer & Brooks, 2011; Hirschmüller, Egloff, Nestler, & Back, 2013; Kenny, Albright, & Malloy, 1988; Kenny, Horner, Kashy, & Chu, 1994). The search for an explanation of this replicable level of accuracy is really a search for what constitutes good information—clearly, some relevant information is present almost immediately[4]. In the studies cited

[4] This may even occur in the form of simple markers of group membership (e.g., gender). For more on stereotype accuracy, see Ch. 16 by Jussim in this handbook.

thus far, participants were generally able to see each other move and hear each other talk[5] (if only briefly). In some cases, there may have even been opportunity for incidental contact surrounding the experimental session. Thus, researchers began to strip all information away, adding bits at a time in order to determine how we become more accurate in judging personality.

In a landmark study in this area, Borkenau and Liebler (1992)[6] asked participants to view targets in one of four conditions. Each target had been photographed and then filmed while walking into a room, sitting down, reading a weather report, standing up, and leaving the room. One group of judges was allowed to see only the photograph of the target individual. Another group only heard the speaking portion of the film. A third group saw the film with no sound, and a fourth group saw the film with audio. One could consider this a (quasi) escalating amount of information available across observers, which should predict greater accuracy (consensus and self-other agreement) in the sound-film group than the silent film group and so forth. Indeed, there was a general trend towards greater accuracy with greater information. However, it is noteworthy that even a still photograph led to greater-than-chance

[5] A person's physiognomy or actions are not the only possible cues to personality; there is an extensive literature covering other types of zero acquaintance research—particularly that which involves no contact with or direct observation of the individual. Given the ubiquity of social media in our time, there have been several studies demonstrating that people's websites (Vazire & Gosling, 2004), social networking profiles and activity (Back et al., 2010; Ivcevic & Ambady, 2012; Tskhay, & Rule, 2014), and even email addresses (Back, Schmukle, & Egloff, 2008) provide valid cues to people's personalities. Additionally, personal artifacts can also be useful in understanding others. A look at one's office or bedroom can generate some accurate trait judgments (Gosling, Ko, Mannarelli, & Morris, 2002), as can listening to someone's playlist or music collection (Rentfrow & Gosling, 2003). There is even work examining the connection between personal odor and personality (Sorokowska, Sorokowski, & Szmajke, 2012) and even what can be gleaned from looking at someone's shoes (Gillath, Bahns, & Crandall, 2012). Thankfully, these two streams of research have not met. For reviews of what can be learned from personal artifacts and spaces and online presence, see Chapter 14 by Wall and Campbell in this handbook. To learn more about what can be gleaned from nonverbal behavior in low acquaintance settings, please see Chapter 13 by Breil, Hirschmüller, Nestler, and Back in this handbook. Finally, Connelly and Ones's (2010) meta-analysis also contains a substantial and interesting component addressing accuracy of zero acquaintance judgments made from different types of information.

[6] Aside from being one of the first studies to systematically control the amount of information exchanged—essentially by removing actual interaction among participants—Borkenau and Liebler (1992) were also among the first to employ Brunswik's (1956) lens model as a means to understanding the mechanism of judgmental accuracy in personality perception. For a description of this model and myriad findings related to it with respect to accuracy in personality judgment, see Chapter 4 by Hirschmüller, Breil, Nestler, and Back this handbook.

levels of self-other agreement for Extraversion in this sample—a finding replicated many times over (Beer, 2013; Beer, 2014; Beer & Watson, 2010; Naumann, Vazire, Rentfrow, & Gosling, 2009), even with exposure to photographs as brief as 50 milliseconds (Borkenau, Brecke, Möttig, & Paelecke, 2009).

Studies of this sort—in which information is carefully manipulated and personality accuracy measured--are actually rather uncommon in the literature. In an early systematic evaluation of differing types and levels of information on personality judgments, Weiss (1979) did not observe gains in consensus, though there was evidence of differential utilization of information across conditions. Further, as mentioned earlier, some models of personality accuracy (Kenny, 1991) would indeed predict that consensus may not increase with more information unless this additional information was shared across judges—judges exposed to different sets of new information about a target may not agree more with one another about that target. Other studies, however, have replicated general beneficial effects of increased quantity of information. For instance, Beer and Watson (2010) replicated the gains from still photographs to videos, demonstrating the utility of dynamic visual and audio cues (beyond static visual cues) in assessing Extraversion, and another group observed increases in accuracy (for some traits) upon a second live interaction of a different kind than the first (i.e., a negotiation task following an unstructured conversation with the same individual; Wall, Taylor, Campbell, Heim, & Richardson, 2018). In a more ambitious study, Borkenau, Mauer, Riemann, Spinath, and Angleitner (2004) presented videos of targets engaging in 15 different activities (e.g., building a paper tower, introducing someone to someone else, mock persuasion, singing) to different judges and found that aggregating impressions from multiple observations was associated with accuracy.[7] Although aggregating information in this fashion is not exactly equivalent to comparing evaluations made by a single individual with greater

[7] This aggregation is beneficial only to a certain point—it seems that any more than six instances yielded little gain in accuracy of aggregated perceptions.

exposure to a target, it is certainly an interesting approximation of the phenomenon. Furthermore, this study shares a design complication with the others discussed in preceding paragraphs: frequently in studies of this nature—as in more naturalistic studies of acquaintanceship effects--amount of information is confounded with type of information. In Borkenau and Liebler's (1992) study, for example, one can be sure that the sound film condition contains more information than the still photo condition, but the distinction between the audio-only and still photo condition is not solely an issue of amount.

Pure Quantity of Information

One area of study in which researchers are clearly focused on amount of information is in *thin slice* paradigms (Ambady, Bernieri, & Richeson, 2000; Ambady & Rosenthal, 1992). Typically, in this kind of work judges are exposed to small samples of verbal and/or nonverbal behavior taken from a larger behavioral observation. For example, researchers might present one or five seconds of a silent recorded lecture by a professor and use this to predict end-of-semester teaching evaluations (e.g., Ambady & Rosenthal, 1993). These thin slices have shown to be useful in predicting various social judgments including things such as sexual orientation (Ambady, Hallahan, & Conner, 1992), nature of relationships (Ambady & Gray, 1995), and intelligence (Murphy, Hall, & Colvin, 2003). This methodology has also been applied specifically to personality judgments. Though an early review indicated that gains in predictive accuracy in general are modest to non-existent from exposures between 30 seconds and 5 minutes (Ambady & Rosenthal, 1992), more recent research indicates that small increases in exposure to the same class of information (usually a video recording of the target interacting with someone) can lead to small gains in accuracy in the domain of personality judgment. For example, Blackman and Funder (1998) observed gains in self-other agreement (but not consensus) when participants viewed 5-10-minute segments of target behavior versus 25-30-minute segments, and Carney, Colvin, and Hall (2007) noted a similar linear trend (though only for certain traits) when participants viewed clips ranging

from 5 seconds to 5 minutes. Letzring, Wells, and Funder (2006) also demonstrated some increases in profile accuracy from initial acquaintance to 50 minutes of conversation (but not from 50 minutes to 3 hours). Finally, Krzyzaniak, Colman, Letzring, McDonald, and Biesanz (2018) observed a linear increase in accuracy for Extraversion and a non-linear increase for Conscientiousness from 30 seconds to 5 minutes of exposure to a target. In short, there is at least some evidence for the positive impact of escalating quantity of information in both shorter- and longer-term observations of the phenomenon.

Pure Quality of Information

Increasing quantity of information while holding quality relatively steady is a simpler feat than altering the quality of information available to a judge without affecting quantity of information exchanged. Nonetheless, researchers have made a few attempts to tackle this latter issue. The first clear attempt to do so involved exposure to video recordings of target individuals discussing either (a) their thoughts and feelings with respect to range of topics (e.g., work, family, important life choices) or (b) their specific behaviors with respect to those same topics (Andersen, 1984). In keeping with lay expectations observed in another study (Andersen & Ross, 1984), the thoughts and feelings videos generated greater accuracy than did the behavioral descriptions. More recently, Letzring and Human (2014) approached information quality from a similar vantage point, pitting thoughts and feelings and discussions of behaviors against actually-performed behaviors (e.g., read a poem aloud, interpret a Thematic Apperception Test card, playing games, explaining idioms)--except in this case the researchers assigned dyads to discuss these things in person as opposed to observing someone in a video. They found that distinctive accuracy (the ability to judge how people differ from the average person) was generally greater when participants discussed thoughts and feelings or discussed behaviors (specific or general) than when they actually engaged in behaviors together. Departing from thoughts and feelings versus behaviors and their descriptions, Beer and Brooks (2011) instead chose another potential parameter of information quality: varying types of self-disclosure. Building on previous work (Pronin,

Fleming, & Steffel, 2008) which established that lay perceivers believe that learning about personal values (e.g., god, family, justice) is more informative than learning distinguishing facts (e.g., can play the trombone, owns a flying squirrel, is afraid of vending machines), participants in small groups were asked to disclose either three core personal values or three distinguishing facts about themselves. Although there was no general advantage (collapsing across trait dimensions) for one type of information over another, those in the facts condition more accurately judged Conscientiousness and those in the values condition more accurately judged Neuroticism. These trait X judgment-context interactions are not uncommon. Wall, Taylor, Dixon, Conchie, and Ellis (2013) evaluated accuracy while escalating "richness" of context (internet chat, telephone, face-to-face conversation) and found greater accuracy for Neuroticism and Extraversion as richness increased, but the reverse pattern for Openness and Conscientiousness. These kinds of findings fuel speculation that relevance (the RAM stage through which information quality primarily operates) needs to be considered in context. For example, one consistent finding in the zero-acquaintance literature is general inaccuracy in assessing Neuroticism. However, this may be explained by the fact that the contexts in which we have examined accuracy at zero acquaintance afford opportunities for behavioral expressions of Extraversion but not Neuroticism. Indeed, Neuroticism was evaluated more accurately in circumstances where the evaluative nature of the social context was particularly salient, which enhances the availability of relevant cues for this trait (Hirschmuller, Egloff, Schmukle, Nestler, & Back, 2015, see also Ch. 8 by Krzyzaniak & Letzring in this handbook).

Joint Evaluations of Quantity and Quality

Studies described in the preceding two sections provide good opportunities to evaluate the impact of quantity and quality in isolation, but some researchers have been able to rather cleverly examine the influence of these factors simultaneously. In a particularly ambitious study, Letzring, Wells, and Funder (2006) constructed five judgment contexts representing increasing quality and quantity of

information. Quantity escalated from a zero-acquaintance judgment situation to a 50-minute unstructured conversation to a 3-hour unstructured conversation. Quality, on the other hand, was represented at the low end by a 50-minute highly-structured interaction (answering trivia questions as a group) and at the high end by a 50-minute conversation with explicit instructions to "get to know" the other participant. Results support that increases in both quantity and quality are associated with greater accuracy. However, in each case the most significant gains occurred from low to medium levels of quantity and quality. To date, this is the only published study in which such systematic evaluations of quantity and quality occur simultaneously.

Evaluation and Future Directions

The preceding review was hardly exhaustive, but rather meant as a primer and roadmap to the various avenues of research with bearing on the question: what is good information? In this final section, I will highlight some of the themes and focus on some of the most important open questions in the subfield. I will also make some suggestions based on these conclusions.

Effect Sizes

One trend in both controlled laboratory studies and more descriptive studies of existing groups is that the effect of information on personality accuracy could be considered rather small[8]. Agreement correlations in laboratory studies may shift from approximately .20 to .30 with additional exposure, and groups that one might imagine are substantially more acquainted (i.e., married couples) may only generate accuracy correlations (correlations typically .50-.60) slightly greater than do relative strangers—at least for Extraversion (correlations typically in the .30s). Biesanz et al. (2007) noted of

[8] The practical impact of an effect can be separated from its statistical effect size, and the desired size of an effect should be considered in context. In this section, I argue as if effects of the size described throughout this chapter are small, though researchers in this field could (convincingly) argue that they are not. See Meyer et al. (2001) for a discussion of such issues.

Watson et al.'s (2000) work that 5 years of acquaintance was associated with an effect size increase of *r* = .05. Allik et al. (2016) observed even smaller gains. Why so modest? Should we be concerned?

One issue, as noted earlier, is that some information available quite early upon encountering a new person is fairly diagnostic, thus the lower bound for accuracy is fairly high. This is particularly true for Extraversion. Similarly, one could also argue that the upper bound is not impressive. Should self-other agreement correlations among individuals married for 30 years not exceed *r* = .60? There are at least two possible explanations to consider for these conditions, both involving measurement and one involving a theoretical shift. First, it is possible that the standard methods for evaluating accuracy are not suited to observing information's effect on accuracy. There have been various debates about what constitutes an accuracy criterion. Clearly, self-judgments are not always accurate (Dunning et al., 2004), and we have already discussed some of the pitfalls of relying solely on consensus of observers as a measure of accuracy. Some researchers have opted to combine methods (e.g., Letzring et al., 2006) in an effort to mitigate the shortcomings of each. This can be effective and is generally recommended, though it does not protect against circumstances in which the component sources share a systematic error component, such as a positivity bias (Leising, Erbs, & Fritz, 2010). Another recommendation is to make more consistent efforts to connect personality judgments to tangible life outcomes and observable behaviors. Some early efforts to compare the accuracy of say, informants versus strangers in predicting laboratory (Vazire, 2010) or natural (Beer & Vazire, 2017) behaviors have yielded mixed results, and these studies were not carefully designed to address the particular issue of information as a moderator. It is possible that predicting concrete outcomes will depress effect sizes across the board but highlight the value of various kinds of information in ways that interjudge agreement does not capture. In sum, more carefully chosen accuracy criteria may yield larger effects.

A second possible explanation for the small effect sizes could be that the upper bound and lower bound accuracy estimates are truncated due to the decontextualized nature of standard

measurements. A frequent complaint of people first encountering most common personality inventories is that the items are too broad to capture their personality fully. What does it mean to be calm and relaxed in a general sense? If one has a history of being calm in emergencies but finds herself frequently and easily agitated by fellow motorists, how does one respond to the item on a 5-point scale of Strongly Disagree to Strongly Agree? There have been strong calls for contextualizing personality in theory and empirical practice (Mischel, 1973; Shoda & Mischel, 1993; Shoda & Mischel, 1996), and it is possible that well-acquainted individuals might be more accurate in evaluating contextualized tendencies rather than general tendencies (e.g., Friesen & Kammrath, 2011). Perhaps accuracy could be better defined as the range of contexts in which one can predict how another person behaves, thinks, or responds emotionally.

Scarcity of Studies

Although there exists a large amount of data regarding accuracy of personality judgments, the number of studies explicitly dedicated to understanding the impact of information on judgmental accuracy is actually rather small. By this researcher's count, there have been fewer than 20 longitudinal investigations into accurate personality judgment, and these typically do not exceed approximately 3 months in length. Laboratory studies in which the quantity of information is increased and accuracy is observed are limited to the few outlined in this chapter, and in these there is a notable absence of within-subject designs (e.g., repeated assessments of the same target by the same judge in response to increasing information). Finally, the formal study of information quality as a moderator is in its infancy—in fact, this review *is* exhaustive in that particular case.

Why has progress been so slow on these issues? There are certainly practical concerns. Longitudinal studies of this sort are difficult and expensive to execute, and the laboratory protocols for repeated assessments can be tedious and taxing for participants. But another reason could be that the

field is still searching for a unified theory to serve as the backdrop for these explorations. Information quality, in particular, has long suffered from a lack of theoretical grounding. For a long time, the simple distinction between strong situations (wherein the situation constrains individuals' behaviors) and weak situations (for an extended discussion, see Cooper & Withey, 2009) could be considered as one of the only theories of what constitutes high quality information—that which is obtained in the context of a "weaker" situation (i.e., a cocktail party, as opposed to waiting in line at the post office). The studies aimed at evaluating the impact of information quality generally espoused some fairly arbitrarily chosen parameter (e.g., thoughts and feelings versus behaviors, distinguishing facts versus personal values, structured versus unstructured activities) and compared accuracy estimates. The field would benefit if researchers had some guiding principles as to how to manipulate quality in such studies.

In addition, the more descriptive studies of information quantity and quality could also benefit from using theory to guide the choice of comparison groups. Like the laboratory work, researchers seemed to simply grab at groups that may differ in amount of information exchanged (e.g., married couples versus dating couples, fraternity seniors versus Peace Corps trainees) without particular attention to why exactly these groups differ. Here I would like to make a case for Starzyk et al.'s (2006) work on clarifying acquaintanceship as a means to guide our empirical efforts. Figure 2 contains a visual representation of information quality versus quantity across various types of relationships as derived from Starzyk et al.'s (2006) model. From this schematic, one could determine that Watson et al.'s (2000) comparison of dating couples versus friendship dyads may be considered as indicative of differences in information quality alone, whereas the comparison between married couples and these groups involves a distinction in both quality and quantity. Of course, actually making that claim would require assessing a given relationship on indices of quantity and quality, rather than broadly comparing types of relationships. In addition, future studies could involve choosing groups that systematically differ on just one dimension at a time in an effort to evaluate Connelly and Ones's (2010) suggestion that differences

in accuracy among acquainted individuals are largely related to quality of information. Aside from that, simply measuring acquaintance systematically may elucidate some of the murkiness in the field. For example, perhaps the aforementioned small effect sizes—particularly in longitudinal studies—may have to do with a fundamental failure of the manipulation. Perhaps roommates or discussion groups simply do not see increases in the quality of their interactions in the form of greater self-disclosure, knowledge of goals, physical intimacy, or knowledge of social network. Quantity is increasing, but not quality (Figure 2 also illustrates this occasional orthogonality). Laboratory studies should involve specific attempts to increase some of these parameters (probably not physical intimacy) as fundamental ingredients in studies of information quality. Even studies of information quantity may benefit from some of the lessons from Starzyk et al. (2006) in that frequency of contact and duration of relationship are often confounded. Is greater accuracy achieved when one interacts with another person routinely for short periods of time (e.g., my interactions with the administrative office workers for the neighboring department) versus more rarely but for longer periods of time (e.g., my interactions with one of our book publisher representatives)? Does the nature of the quantity impact the quality of these interactions, such that longer interactions breed greater self-disclosure? In any case, there is a need for theory development in understanding information as a moderator of judgmental accuracy, whether it be greater attention to work like Staryzk et al. (2006) or development of new integrative frameworks. Doing so may help generate more meaningful studies and greater interest in the field among researchers.

Future Directions

I have already discussed some of the hopes I have for the future of this area of study, but I can make some more specific suggestions for empirical inquiry. The ideal descriptive study is one with a true zero-acquaintance point, involving representative samples of community-dwelling adults, spanning multiple years with multiple measurement occasions in real relationships that are not created for the

purpose of the study. This is practically impossible. The most likely avenue to research of this sort would involve partnering with a dating or matchmaking website, but this would still limit the scope of understanding to the context of romantic relationships. Short of accomplishing all of these goals, researchers should strive for as many as possible in a given study. Why do each of these desired components matter? First, establishing a baseline is difficult, as some information is obtained immediately upon visual contact with an individual, and a fair amount of information can be obtained in just a few minutes of exposure. Second, many purposeful studies of acquaintanceship (particularly longitudinal studies) involve college students, whom we know do not perfectly represent the population to which we would like to generalize (Henrich, Heine, & Norenzayan, 2010). Third, the time period must be long (probably measured in years rather than weeks or months) given what is likely slow growth in accuracy over time (Biesanz et al., 2007). The measurement occasions must be fairly frequent and, perhaps as importantly, they must involve explicit measures of information quantity and quality. Finally, these must be real relationships to help ensure that personality assessments are meaningful to the interactants and the task is taken seriously. Motivation has been shown to increase accuracy (Biesanz & Human, 2010), and lack of motivation on the part of participants cannot be ruled out as a potential cause for the absence of judgmental accuracy in some research paradigms.

For laboratory work, as discussed above, it is important that researchers have good scientific reasons for the chosen manipulations and are mindful of the implications for quantity and quality simultaneously. The field needs more systematic evaluations involving graduated exposure to information about targets. It may also behoove the field to consider reuniting to some extent with the social cognition and judgment and decision-making literature. In the 1980s, the study of accuracy and the study of error or bias in social judgment parted ways theoretically and methodologically (Funder, 1987), and in some ways, the two literatures have carried on somewhat independently over time. But each could still inform the other, as the study of how to be right is not always entirely separate from the

study of how to be wrong. Certainly, researchers in personality have bridged this gap (Luo & Snider, 2008; Rogers & Biesanz, 2014; West & Kenny, 2011; Zimmerman, Schindler, Klaus, & Leising, 2018), but there are more opportunities available. For example, the concept of dilution effects—when lower quality information mitigates or overwhelms higher quality information—have been established in other fields (Nisbett, Zukier, & Lemley, 1981), but not yet in the study of accurate personality judgment (but see Beer, Rogers, & Letzring, 2018).

Conclusion

In sum, there is evidence that we come to know others better over time. This evidence may not be fully consistent or statistically overwhelming, but it does exist. We have some clues as to how it happens, and a few clues as to when. We know that all information is not equally useful, but we are not entirely certain which kinds are most useful. For a field that has existed for approximately a century, we seem to still be scratching the surface to some extent. This is unfortunate for a couple of reasons. First, of the primary moderators originally enumerated by Funder (1995), some have suggested that the information moderator has the most promise for theoretical and empirical development. Allik et al. (2016) argued that the relative ease with which people can achieve judgmental accuracy in personality (contrary to Funder's (1995) claim) renders the search for evidence of good targets, good judges, and good traits generally futile, and that information (quality of information, in particular—in accordance with Connelly & Ones' (2010) assertion) is likely to yield the most useful and interesting avenues of research into moderators of accuracy. Second, the potential real-world implications of understanding information quality and quantity's impact on accuracy could represent a significant advance for personality and social psychology's standing as a practical science. Perhaps we could make better hires and have fewer divorces if we had clearer understanding of exactly how and when we can know exactly what about a person. Thus, perhaps a final suggestion I would make to researchers of these phenomena is to find more ways to clearly connect accuracy to these important real-life outcomes. Perhaps in

identifying the optimal judgment conditions for producing contextualized behavioral predictions in meaningful real-world circumstances (e.g., demonstrating that certain structured interactions can allow us to determine whether someone is likely to cheat when in competition or lash out in anger when challenged), we can attract greater—and deserved, in this author's opinion—attention to this important area of inquiry.

References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, *55*, 387-395. doi:10.1037/0022-3514.55.3.387

Allik, J., de Vries, R. E., & Realo, A. (2016). Why are moderators of self-other agreement difficult to establish? *Journal of Research in Personality*, *63,* 72-83. doi:10.1016/j.jrp.2016.05.013

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna, M. P. Zanna (Eds.), *Advances in experimental social psychology, Vol. 32* (pp. 201-271). San Diego, CA, US: Academic Press. doi:10.1016/S0065-2601(00)80006-4

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. doi:10.1037/0022-3514.69.3.518

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, *77*, 538-547. doi:10.1037/0022-3514.77.3.538

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256-274. doi:10.1037/0033-2909.111.2.256

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, *64*, 431-441. doi:10.1037/0022-3514.64.3.431

Back, M. D., & Vazire, S. (2012). Knowing our personality. In S. Vazire, T. D. Wilson, S. Vazire, T. D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 131-156). New York, NY, US: Guilford Press.

Andersen, 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. doi:10.1037/0022-3514.46.2.294

Andersen, S. M., & Ross, L. (1984). Self-knowledge and social inference: I. The impact of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology*, *46*, 280-293. doi:10.1037/0022-3514.46.2.280

Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality*, *42*, 1116-1122. doi:10.1016/j.jrp.2008.02.001

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*, 372-374. doi:10.1177/0956797609360756

Beer, A. (2014). Comparative personality judgments: Replication and extension of robust findings in personality perception using an alternative method. *Journal of Personality Assessment*, *96*, 610-618. doi:10.1080/00223891.2013.870571

Beer, A. (2013). Group personality judgments at zero acquaintance: Communication among judges versus aggregation of independent evaluations. *Journal of Research in Personality*, *47*, 385-389. doi:10.1016/j.jrp.2013.03.008

Beer, A., & Brooks, C. (2011). Information quality in personality judgment: The value of personal disclosure. *Journal of Research in Personality*, *45*, 175-185. doi:10.1016/j.jrp.2011.01.001

Beer, A., Rogers, K. H., & Letzring, T. D. (2018). Effects of Escalated Exposure to Information on Accuracy of Personality Judgment. Manuscript submitted for publication.

Beer, A., & Vazire, S. (2017). Evaluating the predictive validity of personality trait judgments using a naturalistic behavioral criterion: A preliminary test of the self-other knowledge asymmetry model. *Journal of Research in Personality*, *70,* 107-121. doi:10.1016/j.jrp.2017.06.004

Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment*, *90*, 250-260. doi:10.1080/00223890701884970

Beer, A., & Watson, D. (2010). The effects of information and exposure on self-other agreement. *Journal of Research in Personality*, *44*, 38-45. doi:10.1016/j.jrp.2009.10.002

Berscheid, E., Snyder, M., & Omoto, A. M. (1989). The Relationship Closeness Inventory: Assessing the closeness of interpersonal relationships. *Journal of Personality and Social Psychology*, *57*, 792-807. doi:10.1037/0022-3514.57.5.792

Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions: Accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science*, *21*, 589-594. doi:10.1177/0956797610364121

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, *92*, 119-135. doi:10.1037/0022-3514.92.1.119

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. doi:10.1006/jesp.1997.1347

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*, 703-706. doi:10.1016/j.jrp.2009.03.007

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*, 645-657. doi:10.1037/0022-3514.62.4.645

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. doi:10.1037/0022-3514.86.4.599

Brown, J. A., & Bernieri, F. (2017). Trait perception accuracy and acquaintance within groups: Tracking accuracy development. *Personality and Social Psychology Bulletin*, *43*, 716-728. doi:10.1177/0146167217695557

Brunswik, E. (1956). *Perception and the representative design of psychological experiments., 2nd ed*. Berkeley, CA, US: University of California Press.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*, 1054-1072. doi:10.1016/j.jrp.2007.01.004

Cleeton, G. U., & Knight, F. B. (1924). Validity of character judgments based on external criteria. *Journal of Applied Psychology*, *8*, 215-231. doi:10.1037/h0072525

Cloyd, L. (1977). Effect of acquaintanceship on accuracy of person perception. *Perceptual and Motor Skills*, *44*, 819-826. doi:10.2466/pms.1977.44.3.819

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*, 1092-1122. doi:10.1037/a0021212

Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review*, *13*, 62-72. doi:10.1177/1088868308329378

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*, *5*, 69-106. doi:10.1111/j.1529-1006.2004.00018.x

Estes, S. G. (1938). Judging personality from expressive behavior. *The Journal of Abnormal and Social Psychology*, *33*, 217-236. doi:10.1037/h0058565

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., & Delgado, M. R. (2012). Social network modulation of reward-related signals. *The Journal of Neuroscience*, *32*, 9045–9052. doi.org/10.1523/JNEUROSCI.0610-12.2012

Friesen, C. A., & Kammrath, L. K. (2011). What it pays to know about a close other: The value of if-then personality knowledge in close relationships. *Psychological Science*, *22*, 567-571. doi:10.1177/0956797611405676

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, *101*, 75-90. doi:10.1037/0033-2909.101.1.75

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652-670. doi:10.1037/0033-295X.102.4.652

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*, 177-182. doi:10.1177/0963721412445309

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. doi:10.1037/0022-3514.55.1.149

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. doi:10.1037/0022-3514.69.4.656

Gillath, O., Bahns, A. J., Ge, F., & Crandall, C. S. (2012). Shoes as a source of first impressions. *Journal of Research in Personality*, *46*, 423-430. doi:10.1016/j.jrp.2012.04.003

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*, 379-398. doi:10.1037/0022-3514.82.3.379

Gros, D. F., Simms, L. J., & Antony, M. M. (2010). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) in friendship dyads. *Behavior Therapy*, *41*, 277–284. doi.org/10.1016/j.beth.2009.07.001

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61-83. doi:10.1017/S0140525X0999152X

Hirschmüller, S., Egloff, B., Nestler, S., & Back, M. D. (2013). The dual lens model: A comprehensive framework for understanding self–other agreement of personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, *104*, 335-353. doi:10.1037/a0030383

Human, L. J., & Biesanz, J. C. (2011). Target adjustment and self-other agreement: Utilizing trait observability to disentangle judgeability and self-knowledge. *Journal of Personality and Social Psychology*, *101*, 202-216. doi:10.1037/a0023782

Human, L. J., & Biesanz, J. C. (2012). Accuracy and assumed similarity in first impressions of personality: Differing associations at different levels of analysis. *Journal of Research in Personality*, *46*, 106-110. doi:10.1016/j.jrp.2011.10.002

Human, L. J., & Biesanz, J. C. (2013). Targeting the good target: An integrative review of the characteristics and consequences of being accurately perceived. *Personality and Social Psychology Review*, *17*, 248-272. doi:10.1177/1088868313495593

Ivcevic, Z., & Ambady, N. (2012). Personality impressions from identity claims on Facebook. *Psychology of Popular Media Culture*, *1*, 38-45. doi:10.1037/a0027329

John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, O. P. John, L. A. Pervin, O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). New York, NY, US: Guilford Press.

Kahn, J. H., Hucke, B. E., Bradley, A. M., Glinski, A. J., & Malak, B. L. (2012). The Distress Disclosure Index: A research review and multitrait–multimethod examination. *Journal of Counseling Psychology*, *59*, 134–149. doi.org/10.1037/a0025716

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, *98*, 155-163. doi:10.1037/0033-295X.98.2.155

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. doi:10.1037/0033-2909.116.2.245

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, *62*, 88-97. doi:10.1037/0022-3514.62.1.88

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self- and other perception: A meta-analysis. *Personality and Social Psychology Review*, *14*, 196-213. doi:10.1177/1088868309353414

Krzyzaniak, S., Colman, D. C., Letzring, T. D., McDonald, J. S., & Biesanz, J. C. (2018) *The effects of information quantity on the distinctive and normative accuracy of personality trait judgments.* Manuscript submitted for publication.

Kurtz, J. E., & Sherker, J. L. (2003). Relationship quality, trait similarity, and self-other agreement on personality ratings in college roommates. *Journal of Personality*, *71*, 21-48. doi:10.1111/1467-6494.t01-1-00005

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality*, *42*, 914-932. doi:10.1016/j.jrp.2007.12.003

Letzring, T. D. (2015). Observer judgmental accuracy of personality: Benefits related to being a good (normative) judge. *Journal of Research in Personality*, *54*51-60. doi:10.1016/j.jrp.2014.05.001

Letzring, T. D., & Human, L. J. (2014). An examination of information quality as a moderator of accurate personality judgment. *Journal of Personality*, *82*, 440-451. doi:10.1111/jopy.12075

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, *91*, 111-123. doi:10.1037/0022-3514.91.1.111

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*, 668-682. doi:10.1037/a0018771

Luo, S., & Snider, A. G. (2009). Accuracy and biases in newlyweds' perceptions of each other: Not mutually exclusive but mutually beneficial. *Psychological Science*, *20*, 1332-1339. doi:10.1111/j.1467-9280.2009.02449.x

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does 'failure to replicate' really mean? *American Psychologist*, *70*, 487-498. doi:10.1037/a0039400

McCrae, R. R. (1994). The counterpoint of personality assessment: Self-reports and observer ratings. *Assessment*, *1*, 159-172. doi:10.1177/1073191194001002006

McCrae, R. R., Stone, S. V., Fagan, P. J., & Costa, P. J. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality*, *66*, 285-313. doi:10.1111/1467-6494.00013

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*, 128–165. doi.org/10.1037/0003-066X.56.2.128

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*, 252-283. doi:10.1037/h0035002

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators and gender effects. *Journal of Personality*, *71*, 465-493. doi:10.1111/1467-6494.7103008

Nauta, M. M. (2012). Are RIASEC interests traits? Evidence based on self–other agreement. *Journal of Career Assessment*, *20*, 426–439. doi.org/10.1177/1069072712448895

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, *35*, 1661-1671. doi:10.1177/0146167209346309

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, *13*, 248-277. doi:10.1016/0010-0285(81)90010-4

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, *4*, 681-691. doi:10.1037/h0024002

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. doi:10.1037/0022-3514.63.5.816

Paulhus, D. L., & Reynolds, S. (1995). Enhancing target variance in personality impressions: Highlighting the person in person perception. *Journal of Personality and Social Psychology*, *69*, 1233-1242. doi:10.1037/0022-3514.69.6.1233

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. doi:10.1037/0022-3514.72.3.604

Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of Personality and Social Psychology*, *56*, 823-833. doi:10.1037/0022-3514.56.5.823

Pronin, E., Fleming, J. J., & Steffel, M. (2008). Value revelations: Disclosure is in the eye of the beholder. *Journal of Personality and Social Psychology*, *95*, 795-809. doi:10.1037/a0012710

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*, 1236-1256. doi:10.1037/0022-3514.84.6.1236

Rogers, K. H., & Biesanz, J. C. (2014). The accuracy and bias of interpersonal perceptions in intergroup interactions. *Social Psychological and Personality Science*, *5*, 918-926. doi:10.1177/1948550614537307

Rogers, K. H., & Biesanz, J. C. (2018). Reassessing the good judge of personality. *Journal of Personality and Social Psychology*, doi:10.1037/pspp0000197

Schneider, L., Schimmack, U., Petrican, R., & Walker, S. (2010). Acquaintanceship length as a moderator of self-informant agreement in life-satisfaction ratings. *Journal of Research in Personality*, *44*, 146-150. doi:10.1016/j.jrp.2009.11.004

Shen, E. (1925). The influence of friendship upon personal ratings. *Journal of Applied Psychology*, *9*, 66-68. doi:10.1037/h0075606

Shoda, Y., Mischel, W., & Wright, J. C. (1993). Links between personality judgments and contextualized behavior patterns: Situation-behavior profiles of personality prototypes. *Social Cognition*, *11*, 399-429. doi:10.1521/soco.1993.11.4.399

Shoda, Y., & Mischel, W. (1996). Toward a unified, intra-individual dynamic conception of personality. *Journal of Research in Personality*, *30*, 414-428. doi:10.1006/jrpe.1996.0029

Sorokowska, A., Sorokowski, P., & Szmajke, A. (2012). Does personality smell? Accuracy of personality assessments based on body odour. *European Journal of Personality*, *26*, 496-503. doi:10.1002/per.848

Sparling, S., & Cramer, K. (2015). Choosing the danger we think we know: Men and women's faulty perceptions of sexually transmitted infection risk with familiar and unfamiliar new partners. *Canadian Journal of Human Sexuality*, *24*, 237–242. doi.org/10.3138/cjhs.243-A2

Starzyk, K. B., Holden, R. R., Fabrigar, L. R., & MacDonald, T. K. (2006). The personal acquaintance measure: A tool for appraising one's acquaintance with any person. *Journal 0f Personality and Social Psychology*, *90*, 833-847. doi:10.1037/0022-3514.90.5.833

Taft, R. (1966). Accuracy of empathic judgments of acquaintances and strangers. *Journal of Personality and Social Psychology*, *3*, 600-604. doi:10.1037/h0023288

Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality*, *49*25-30. doi:10.1016/j.jrp.2013.12.004

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*, 281-300. doi:10.1037/a0017908

Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality Impressions Based on Personal Websites. *Journal of Personality and Social Psychology*, *87*, 123-132. doi:10.1037/0022-3514.87.1.123

Wall, H. J., Taylor, P. J., Campbell, C., Heim, D., & Richardson, B. (2018). Looking at the same interaction and seeing something different: The role of informational contexts, judgment perspective, and behavioral coding on judgment accuracy. *Journal of Individual Differences*, *39*, 123-141. doi:10.1027/1614-0001/a000257

Wall, H. J., Taylor, P. J., Dixon, J., Conchie, S. M., & Ellis, D. A. (2013). Rich contexts do not always enrich the accuracy of personality judgments. *Journal of Experimental Social Psychology*, *49*, 1190-1195. doi:10.1016/j.jesp.2013.05.010

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, *57*, 120-128. doi:10.1037/0022-3514.57.1.120

Watson, D., & Clark, L. A. (1991). Self- versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology*, *60*, 927-940. doi:10.1037/0022-3514.60.6.927

Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*, 546-558. doi:10.1037/0022-3514.78.3.546

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, *118*, 357-378. doi:10.1037/a0022936

Zimmermann, J., Schindler, S., Klaus, G., & Leising, D. (2018). The effect of dislike on accuracy and bias in person perception. *Social Psychological and Personality Science*, *9*, 80-88. doi:10.1177/1948550617703167

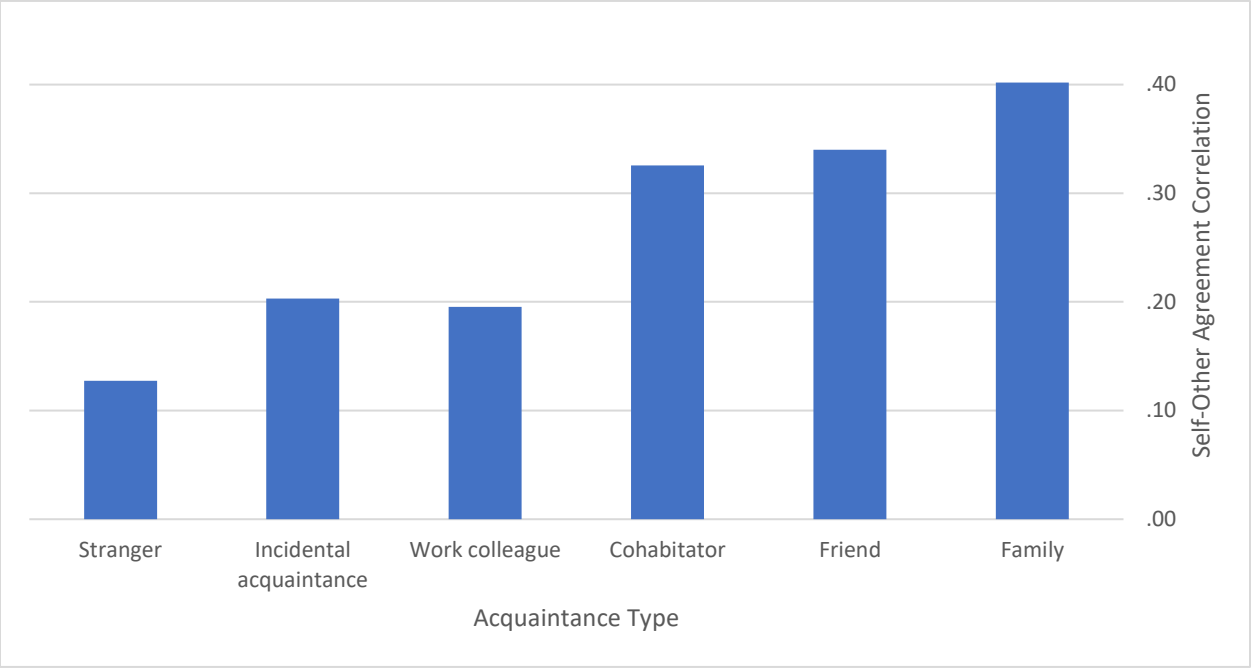

Figure 1. Average self-other agreement correlations (calculated from Connelly & Ones, 2010).

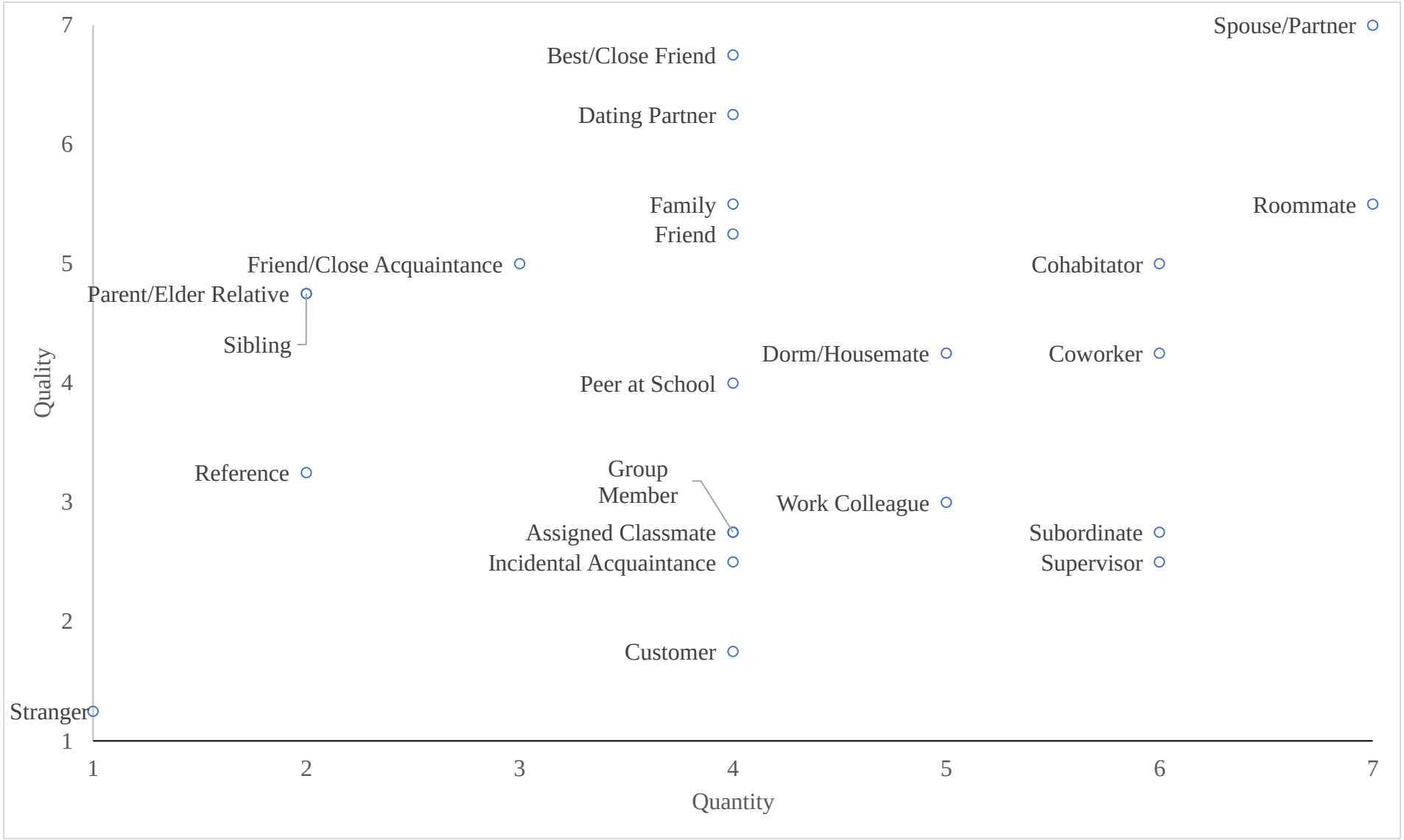
Spouse/Partner
Best/Close Friend
Dating Partner
Roommate
Family
Friend
Friend/Close Acquaintance
Cohabitator
Parent/Elder Relative
Sibling
Dorm/Housemate
Coworker
Peer at School
Reference
Group Member
Work Colleague
Assigned Classmate
Subordinate
Incidental Acquaintance
Supervisor
Customer
Stranger
Quality
Quantity
1
2
3
4
5
6
7

Figure 2. Information Quantity versus Quality across Relationship Types

To create this figure, I quantified the distinctions contained in Table 1 of Connelly & Ones (2010). They asked independent coders to evaluate each type of relationship along the six dimensions of acquaintance established in Starzyk et al. (2006) on a 7-point scale (1 = Very Low, 7 = Very High). To ease presentation, I used frequency of contact as the quantity domain and aggregated the quality-relevant indices into a single quality index.