

Running Head: CONTEXT AND PERSONALITY JUDGMENT

Research Report:

**Rich Contexts Do Not Always Enrich the Accuracy of Personality Judgments**

**Abstract**

We test the common assumption that information ‘rich’ contexts lead to more accurate personality judgments than information ‘lean’ contexts. Pairs of unacquainted students rendered judgments of one another’s personalities after interacting in one of three, increasingly rich, contexts: Internet ‘chat’, telephone, or face-to-face. Accuracy was assessed by correlating participants’ judgments with a measure of targets’ personalities that averaged self and informant ratings. As predicted, the visible traits of extraversion and conscientiousness were judged more accurately than the less visible traits of neuroticism and openness. However, judgment accuracy also depended on context. Judgments of extraversion and neuroticism improved as context richness increased (i.e., from Internet ‘chat’ to face-to-face), whereas judgments of conscientiousness and openness improved as context richness decreased (i.e., from face-to-face to Internet ‘chat’). Our findings suggest that context richness shapes not only the availability of personality cues but also the relevance of cues in any given context.

Keywords: personality judgments; trait visibility; accuracy; context richness, cue availability; first impressions

### **Rich Contexts Do Not Always Enrich the Accuracy of Personality Judgments**

Olivia and Peter were shortlisted to interview for a PhD place at a University that sought students who were thoughtful and conscientious. While Olivia was interviewed face-to-face, Peter's interview was conducted over email to avoid him having to travel internationally. Both candidates were asked the same interview questions. A week later, Peter was offered the position, as the interview panel judged him to be more thoughtful and disciplined. Olivia found this surprising as she viewed herself as very conscientious, and her friends often said the same. If Olivia had been interviewed over email, would that have made a difference?

The question of how, when, and why context affects our ability to judge another's personality has only begun to be addressed in the literature. Most existing studies compare judgments made under various conditions in the same context, such as face-to-face interactions (Carney, Colvin, & Hall, 2007) or Internet 'chat' (Gosling, Gaddis, & Vasire, 2007). These studies typically show that, other things being equal, more information leads to more accurate judgments. For example, individuals presented with verbal and nonverbal stimuli tend to render more accurate judgments about a target's conscientiousness and emotional stability compared to those presented with only visual stimuli (Scherer, Scherer, Hall, & Rosenthal, 1977). Similarly, Letzring, Wells, and Funder (2006) found that increasing the duration of exposure to a target, or increasing the relevance of the content of the exposure, improved judgment accuracy. To be sure, this does not mean that long exposure is necessary for accurate judgments; judgments made after 'thin-slice' exposure to a target can be accurate (Ambady & Rosenthal, 1993). However, even in these circumstances, greater exposure often leads to more accurate predictions (Carney, Colvin, & Hall, 2007; cf. Ambady & Rosenthal, 1993, Study 3).

Interestingly, the positive relationship between information availability and accuracy is

more nuanced when judgments are considered across personality traits. This is particularly evident in studies of ‘trait visibility’, which show that traits associated with frequent and unambiguous behavioral cues (i.e., visible traits) are more accurately judged than traits associated with ambiguous cues (Blackman & Funder, 2002). For example, following a face-to-face interaction, strangers are most accurate at judging their partner on the visible traits of extraversion and, to a lesser extent, conscientiousness (Funder & Colvin, 1997; Watson, 2000). Conversely, judgments of the less visible traits of neuroticism and openness are rarely accurate, even when the judge is able to observe the target over a prolonged period (Shevlin, Walker, Davies, Banyard, & Lewis, 2003).

The strength of evidence linking cue availability to better judgments *within* contexts makes it tempting to extend this finding to hypotheses about judgments *across* contexts. Interaction contexts determine the kinds of behavior that a judge can observe, which may in turn impact on the kinds of cues that are utilized during a judgment. For example, ‘information-rich’ contexts such as face-to-face interactions make available a greater breadth of cues than ‘information-lean’ contexts such as email (Daft & Lengel, 1986). One might hypothesize, then, that the greater richness of cues provided by face-to-face interactions will prove valuable to a judge and that more accurate judgments will be possible as a result. However, findings on judgments made in non-face-to-face contexts are not entirely consistent with this assumption. A number of studies of lean contexts report accuracy levels that are greater than those reported in equivalent face-to-face studies. For example, the information contained within emails has been shown to allow comparatively accurate judgments for the trait of openness (Markey & Wells, 2002; Vasire & Gosling, 2004), even though judgments for openness are difficult to make accurately when based on face-to-face information (Borkenau & Liebler, 1992; Funder &

Dobroth, 1987).

Another striking example of this differentiation of accuracy across contexts relates to people's judgments of extraversion and conscientiousness. These are traditionally viewed as the two most accurately judged traits, with extraversion being the more accurately judged of the two (Albright, Kenny, & Malloy, 1988; Funder & Colvin, 1988). However, this belief is largely based on analyses of judgments made in face-to-face interactions. Studies examining judgments made in contexts devoid of the cues available in face-to-face interactions show the opposite pattern of accuracy. That is, judgments of conscientiousness tend to be *more* accurate than judgments of extraversion when based on appraisals of physical spaces (Gosling, Ko, Mannarelli, & Morris, 2002), email content (Back, Schmukle, & Egloff, 2008), and stream of consciousness writing (Li & Chignell, 2010; Yee, Harris, Jabon, & Bailenson, 2011).

What might explain this differential performance in trait-level judgments across contexts? A likely explanation is that information-lean contexts, owing to the absence of behavioral cues, encourage judges to focus on the content of communication. This idea is explicit in Culnan and Markus' (1987) distinction between 'cues filtered in' and 'cues filtered out' approaches to understanding the impact of communication technology. A 'cues filtered out' approach posits that face-to-face interactions are rich because they offer high verbal and nonverbal content and allow information to be exchanged simultaneously. Without this immediate access to behavior, people find it difficult to express their individuality (Kiesler, 1986), and this leads to deterioration in the information available to a receiver. By contrast, a 'cues filtered in' perspective suggests that information-lean contexts are beneficial because they encourage senders to focus on communication content, and provide extra possibilities for presenting this content, which is not available in information-rich contexts. These differences may lead

communicators to consider cues they neglect in face-to-face interactions (Walther, Slovacek, & Tidwell, 2001). When a trait is exemplified by content cues, this may lead to more accurate judgments in information-lean contexts than in information-rich contexts.

By implication, we might conjecture about context-driven differences in judgment accuracy for other, less visible, traits. Neuroticism, for example, has been shown to be associated with visual, nonverbal, cues such as being ‘healthier looking’ and ‘standing in a relaxed manner’ (Naumann, Vazire, Rentfrow, & Gosling, 2009). Similarly, in Funder and Sneed’s (1993) extensive study, four of the five largest correlations for neuroticism related to behaviors arguably best expressed by nonverbal behavior (e.g., ‘shows signs of anxiety,’ ‘has awkward interpersonal style’). By contrast, Funder and Sneed found that valid indicators of openness related to content-orientated behaviors such as ‘discussing philosophical issues’ and ‘valuing intellectual matters.’ This is consistent with Pennebaker and King’s (1999) finding that openness was associated with the greatest number of linguistic correlates (10, compared to 5 for neuroticism), and research showing that accurate judgments of openness are possible from email (Hancock & Dunham, 2001; Vasire & Gosling, 2004). Thus, in a similar manner to the differences observed for extraversion and conscientiousness, these results suggest that neuroticism may be judged more accurately in information-rich contexts whereas openness may be judged more accurately when content comes to the fore.<sup>1</sup>

### **Current Study**

In this study we explore the possibility that judgment accuracy depends on both context richness and trait visibility. We propose that some traits, such as extraversion and neuroticism, are judged principally on a target’s expressive behavior, and as such they are better judged in face-to-face contexts where the rich behavioral information is available. By contrast, we propose

that other traits, such as conscientiousness and openness, are judged principally on content information, and as such are better judged in information-lean contexts where content is not obscured by behavioral cues. Finally, consistent with previous research, we predict that the absolute accuracy of judgments depends on trait visibility. Specifically, although we conceive extraversion and neuroticism as behaviorally-orientated traits, we expect extraversion to be judged more accurately than neuroticism because of its visibility. An equivalent prediction is made for conscientiousness over openness, because the former is a positive trait that is more likely to be made visible by targets' behaviors.

We test these predictions by comparing judgments of Big-5 personality traits across three increasingly rich contexts: Internet 'chat', telephone, and face-to-face. Internet 'chat' is the least information-rich context because it relies solely on message content. By contrast, face-to-face interaction is the most information-rich context because it provides access to verbal, nonverbal and paralinguistic cues. Telephone exchanges sit somewhere in the middle of these two contexts: they are driven by message content but provide access to some paralinguistic cues.

## **Method**

### **Participants**

A total of 238 participants (Mean age = 23.8 years; Range 18 - 41) were recruited from two Universities in the North West of England. Of the 238 participants, a core group of 92 students (52 females) participated in one of three interaction conditions: Internet 'Chat' ( $n = 31$ )<sup>2</sup>, Telephone ( $n = 30$ ), and Face-to-Face ( $n = 31$ ). They received course credits, or payment of £5, for their participation. The self-reported ethnicity of this core group was 11% Black African, 6% Indian, 4% Italian, 8% White American, 69% White British, and 2% Other.

The remaining 146 participants were Informants (e.g., friends, family) whose contact details were provided by the Interactants following a request to nominate two people “who knew them well.” These Informants were contacted within two weeks of the initial interaction and asked to report on the Interactant’s personality via post (36%) or email (64%). Their responses enabled realistic personality scores to be computed. All Interactants had at least one Informant rating, and 58.7% received an additional rating from a second Informant.

### **Materials**

Interactants and Informants rendered judgments of personality on a 100-item Big-5 questionnaire derived from the International Personality Item Pool (IPIP, Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006). Used in a range of previous studies (e.g., Ashton & Lee, 2005; Christiansen, Wolcott-Burnam, Janovics, Burns, & Quirk, 2005), the IPIP Big-5 measure has been shown to have good construct validity (Buchanan, Johnson, & Goldberg, 2005) and to correlate highly with corresponding NEO-PI-R scores (International Personality Item Pool, 2011). The 100-item measure comprises 5 sets of 20 items measuring extraversion, neuroticism, conscientiousness, openness, and agreeableness.<sup>3</sup> For each item, participants rate on a 5-point scale the extent to which the statement described the person whom they were rating. Example items are “I am the life of the party” (measuring extraversion) and “I get stressed out easily” (measuring neuroticism).

Three versions of the questionnaire, identical in item content but prefaced with different instructions, were created to make the measure relevant for the different participant roles. Specifically, the instruction “Please use the rating scale below to describe how accurately each statement describes” was suffixed with either: i) *you*; ii) *the person you have just interacted with*; or, iii) *the person who has nominated you*, in order to vary the subject of the measure.



The measure's reliability was assessed across Interactants' self-ratings, Informants' ratings of Interactants, and Interactants' ratings of their partner. Cronbach's alpha for the Internet chat, telephone, and face-to-face responses, respectively, were: .94, .93 and .94 for extraversion; .89, .84 and .80 for neuroticism; .95, .92, and .88 for conscientiousness; .90, .86 and .85 for openness; and .89, .90, and .92 for agreeableness. Cronbach alphas averaged across the three contexts were .89 for Interactants' self-ratings, .81 for Informants' ratings, and .89 for Interactants' ratings of their partner.

### **Procedure**

Interactants signed up to an experiment that focused on "Interpersonal perception between strangers." In order to ensure that they remained unacquainted, they were allocated to pairs and scheduled to arrive to separate rooms five minutes apart. On arrival at the laboratory, Interactants received instructions explaining the study's purpose and, after providing consent, they completed a self-rating of personality. They were then allocated an interaction partner and asked to confirm that they had not met that partner previously. Those who had met their interaction partner ( $n = 8$ ) were re-assigned another partner and rescheduled. Interactants then engaged in a 10 minute interaction with their designated partner in which they were instructed to "talk about anything they wanted." The use of 10 minutes is consistent with other studies (Blackman, 2002; Letzring et al., 2006) and sought to provide participants sufficient time to share personal information and talk about multiple topics. The instruction to talk about "anything they wanted" is also consistent with previous studies (Letzring et al., 2006; Markey & Wells, 2002) and sought to create a context that encouraged the expression of individual differences.

The interactions took place in one of three mediums to which participants were randomly allocated: Internet chat, Telephone, or Face-to-face. In the Internet chat condition, Interactants

were instructed on how to use the ‘chat’ software (Internet ‘chat’ was chosen over email to ensure all three mediums involved synchronous exchanges; see also Gosling, Gaddis, & Vasire, 2007). When acquainted with the software, they were left to chat with their designated partner, who was located in another room, until the experimenter returned. In the telephone condition, the experimenter invited the Interactants to dial the number of a telephone situated with their designated partner in another room. The other Interactant was informed that, when the experimenter left the room, the telephone would ring and that they should answer and talk with another participant until the experimenter returned. Half of the participants were instructed to make the call, while the other half were instructed to wait for the call. In order to minimize any inadvertent power manipulation, they were also told that being asked to receive or make the call was determined randomly. Finally, in the Face-to-face condition, Interactants were taken (separately) from their rooms to a video lab, where they sat at a table and conversed until the experimenter returned.

After their conversation, participants were escorted to their allocated room for the ‘other-assessment’ stage. Each Interactant was asked to complete the IPIP Big-5 measure on the person with whom they interacted “honestly and accurately,” to provide the names of two people who they felt “knew them well,” to indicate how long each person had known them, and to give email or postal addresses so that the person could be contacted. (Interactants reported knowing their Informants for an average of 11.6 yrs,  $SD = 9.6$ ).

### **Assessing Judgment Accuracy**

To avoid some of the problems inherent in self-report (e.g., socially desirable responding; Borkeu & Liebler, 1992), we examined accuracy by comparing judges’ ratings against an average of a target’s self-rating and up to two informant ratings.<sup>3</sup> Informant ratings

represent additional sources of personality assessment that can, to some extent, offset the bias that may be associated with a self-report measure (Letzring et al., 2006). The average correlation between Interactants' self-ratings and Informant ratings was .50 (Range .32 - .72), which is comparable with the correlations observed in other research (e.g., Bagby, Rector, Bindseil, Dickens, Levitan, & Kennedy, 1998; Kurtz & Putnam, 2006; Letzring et al., 2006; Wagerman & Funder, 2007).

To examine the impact of context at a trait level, we examined judgment accuracy item-by-item across participants. This approach had the advantage of removing the variance in responding that results from differences in mean responses to items. Such differences can manifest as apparent accuracy when a person's ratings match a target's ratings simply because the person is able to predict the responses of a "generalized other" (Cronbach, 1955, p. 7). However, while avoiding this potential artifact, a limitation of examining accuracy across items is that it can be confounded by a person's tendency to use particular ratings on all items (i.e., elevation) or on subsets of items (i.e., differential elevation; Cronbach, 1955). To control for this, we took our item-by-item analysis a step further by standardizing the responses of participants across the 100 items. This effectively 'centered' each Interactant's responses, such that elevations in rating preference were not interpreted as judgment error within the analysis.

Thus, our analysis involved three stages. First, using *z*-scores, we standardized the responses across the 100-item IPIP for each judge and for each aggregate target personality measure (i.e., average of self and informant ratings). Second, examining the conditions separately, we correlated for each item the standardized judges' ratings on a personality trait with the standardized aggregate target measure. This produced for each condition a total of 100

correlations (i.e., one for each item) that comprised 20 correlations for each trait. Third, these item-level correlations were subjected to analyses across the conditions to test our hypotheses.

It is worth acknowledging that this approach, like any approach taken to address the mathematical artifacts identified by Cronbach (1955), is not without its problems. Our approach sought to be conservative by avoiding the inclusion of artifactual variance, but it may also have removed substantive variance relating to judges' performance. However, critically, our hypotheses are not concerned with absolute accuracy but with relative accuracy across conditions. Thus, while our conservative approach may produce accuracy correlations that fall toward the low end of those observed in previous research (cf. Back et al., 2008; Naumann et al., 2009), it provided a relatively artifact free test of the performances of judges across the conditions.

### **Results**

Before testing the hypotheses, we sought to ensure that there were no qualitative differences in the substance of what was discussed in each condition. If the interactions across conditions differed in their qualitative content—such as their references to topics such as work, health, family, and achievement—then it could be argued that any observed differences between conditions were the result of differences in what was discussed rather than differences in the cues that were available. To test for this possibility, we submitted a random selection of ten transcribed interactions from each of the three contexts to analysis by the software Linguistic Inquiry and Word Count (LIWC, Pennebaker, Booth, & Francis, 2007). Among other things, LIWC enables the analysis of the extent to which texts contain references to seven types of personal concern (i.e., achievement, death, home, leisure, money, religion, work) and six types of psychological process (affective, biological, cognitive, perceptual, social, and relative)(see

Pennebaker, Chung, Ireland, Gonzales, and Booth, 2007, for full details). An examination of these categories across the conditions provided an assessment of the extent to which there were qualitative differences in the content of the interaction. Such use of LIWC has proven useful in previous research (Danescu-Niculescu-Mizil, Gamon, & Dumais, 2011; Pennebaker & King 1999; Yarkoni, 2010).

In line with the recommendations of Huberty and Morris (1989), the thirteen LIWC categories were submitted to a series of one-way ANOVAs with condition as the between-subject factor. A Bonferroni correction of  $\alpha = .0038$  was used to control for the family-wise error rate. There were no significant differences in category occurrence across the conditions, either for the personal concern categories,  $F(2, 27) < 4.89, ns$ , or for the psychological process categories,  $F(2, 27) < 5.91, ns$ .

In our study, Interactants acted as both a target and a judge, which created non-independence in the data. To evaluate the impact of this non-independence, we computed intraclass correlations (ICC) at both the individual and aggregate level (Shrout & Fleiss, 1979) for each Big-5 trait. A one-way random effects model revealed a mean ICC of .13 (Range .02 to .29). As none of the ICC's exceeded .3 individually or in the aggregate, analyses were computed with individual participants as the unit of analysis (see Kenny, 1995, Table 4; Kurtz & Sherker, 2003). The data were submitted to a 3 (Context: Internet 'chat' vs. Phone vs. Face-to-Face) x 2 (Visibility: High trait visibility [extraversion and conscientiousness] vs. Low trait visibility [neuroticism and openness]) x 2 (Relevance: Content cues vs. Behavioral cues) mixed-design ANOVA with accuracy scores as the Dependent Variable.

Figure 1 presents the accuracy scores for extraversion, conscientiousness, neuroticism and openness as a function of judgment context. In support of our hypothesis, there was a

significant main effect of trait Visibility,  $F(1,57) = 16.87, p < .001, \eta^2 = .30$ , with the two visible traits of extraversion and conscientiousness being judged with greater accuracy than the less visible traits of neuroticism and openness. Similarly, in support of our prediction about information relevance, there was a significant Context x Relevance interaction,  $F(1,57) = 5.63, p < .01, \eta^2 = .20$ . The accuracy for the judgments of extraversion and neuroticism increased with context richness, whereas the accuracy of judgments for conscientiousness and openness decreased with context richness (interaction contrast = .21,  $F(1,117) = 5.23, p < .05$ ). The Relevance x Visibility interaction was not significant,  $F < 1$ , suggesting that these two dimensions worked independently to shape judgment accuracy (the three-way interaction was also not significant).

### Discussion

The accuracy of people's judgments of others' personalities is often assumed to be dependent on the richness of information available to them. This assumption derives largely from research on face-to-face acquaintanceship, which typically finds a positive correlation between the time judges have to observe a target and their judgment accuracy (Carney, Colvin, & Hall, 2007; Letzring et al., 2006). This well-established finding has informed our presuppositions about the accuracy of judgments made *across* contexts. Contexts differ in the richness of the cues that they offer a judge and an assumption is often made that a greater breadth or richness of cues is likely to result in better judge performance. Researchers tend to assume, for example, that face-to-face interactions facilitate more accurate personality judgments than interactions conducted by telephone or email. In this article, we have shown that the relationship between context richness and judgment accuracy is more nuanced, such that there is an optimal fit between context and the trait being judged.

Our study has examined the accuracy of judgments across three increasingly rich

interaction contexts (email, telephone, and face-to-face). The findings confirm that judgment accuracy depends on both the type of trait being judged and the context in which a judgment is made. Consistent with previous research (e.g., Albright, Kenny, & Malloy, 1988; Funder & Colvin, 1988), we found that judgment accuracy was influenced by trait visibility. Participants made more accurate judgments on the visible traits of extraversion and conscientiousness than they did on the less visible traits of neuroticism and openness. However, the extent to which this ‘trait visibility’ effect held true varied as a function of judgment context. Judgments of extraversion were associated with a linear increase in accuracy as the richness of the context increased; conversely, judgments of conscientiousness increased in accuracy when the context became less rich. This finding suggests that trait visibility is moderated by context.

Importantly, this same pattern emerged for the more private traits of neuroticism and openness. Judgments of neuroticism, a trait often associated with nonverbal behaviors (Funder & Sneed, 1993), increased in accuracy as the richness of the context increased. For judgments of openness, a trait often associated with content-related behavior, the opposite was true. Thus, whereas the private nature of these less visible traits reduced the absolute accuracy of judgments, their relative accuracy depended on whether or not a context contained both content and behavioral cues, or solely content cues.

An interesting difference between the extraversion-conscientiousness and neuroticism-openness pairs concerns the point at which the relative accuracy of participants’ judgments switches from one trait to the other. For neuroticism and openness this change occurs between telephone and face-to-face interactions (see Figure 1), suggesting that nonverbal behaviors play a particularly important role in the judgment of these two traits. This is consistent with evidence showing that valid cues to neuroticism and openness are almost exclusively nonverbal behaviors.

For example, ‘appearing relaxed and comfortable’, ‘behaving in a timid manner’, and ‘being healthier looking’ are valid cues for neuroticism (Funder & Sneed, 1993; Naumann et al., 2009), while ‘looking away from the camera’ and ‘having a distinctive style of dress’ are valid cues for openness (DeGroot & Gooty, 2009; Naumann et al., 2009). Conversely, the difference in judgments of extraversion and conscientiousness appears between email and telephone interactions, suggesting that paralinguistic cues may be particularly salient to judgments of these traits. This is also consistent with the literature, which shows that valid cues to extraversion include nonverbal and paralinguistic cues such as ‘speaks in a loud voice’ and ‘is talkative’ (Funder & Sneed, 1993), while valid cues to conscientiousness include cues such as speech rate (Smith, Brown, Strong, & Rencher, 1975), use of big words (Yee, Harris, Jabon, & Bailenson, 2011), and ‘taking a conversation to a higher level’ (Wall, 2011).

These observations complement research showing that accurate judgments of personality are possible following only minimal exposure to (i.e., thin slices of) the target’s behavior. It is often the case in thin-slicing research that exposure to cues is constrained by limiting the channel available to judges, such as videos without sound (Ambady & Rosenthal, 1993) or paralinguistic cues without sight of the target (Addington, 1968). It may be the case that judgments in these circumstances are facilitated precisely because judges are being asked to focus on the kind of cues that are relevant to the judgment being made. For example, in Ambady and Rosenthal’s (1993) study, judges used nonverbal cues to make accurate judgments of teacher ratings. In prior research, Murray, Rushton, and Paunonen (1990) found that extraversion was the trait most discriminant of uniformly good and bad teachers. By focusing on nonverbal behavior, Ambady and Rosenthal’s judges based their ratings on behavioral cues that our findings (and those of others, Funder & Sneed, 1993) suggest are particularly relevant for judgments of extraversion.



Thus, by focusing on nonverbal cues, judges were directed to cues relevant to judgments of teaching excellence. Interestingly, even in scenarios where all channels are available, the time pressure of a thin-slice may encourage judges to focus on particular ‘ecologically valid’ cues. It will be interesting to learn more about the optimal fit between cue and context in order to determine whether or not there are individual differences in people’s ability to focus on the appropriate cues.

Overall, the main implication of our findings is that information-rich contexts are not always preferential when judging personality. It appears that cues relevant to judging traits such as conscientiousness and openness may become ‘washed out’ by the paralinguistic and nonverbal cues that form a part of information-rich contexts. This may be explained using the Realistic accuracy Model (RAM) of personality judgment (Funder, 1999). According to RAM, the accuracy of personality judgments is determined by the degree to which cues are available to the judge, providing that such cues are both relevant to the judgment being made and detected and used appropriately by the judge. When relevance, detection and utilization are held constant, RAM predicts that greater availability of information will lead to more accurate personality judgments. However, as RAM would predict, and our data suggests, it is possible for these four stages to interact. In our case, having a greater breadth of information available appeared to decrease judges’ ability to detect the more subtle cues that allow accurate judgments of traits like conscientiousness.

There are, however, two additional factors that may partially explain our findings. Although the instructions to Interactants were identical in each condition (i.e., to talk about whatever they wanted), the contexts themselves may have affected the way in which judges utilized the cues available to them. Our LIWC comparisons established that the three contexts

were comparable in qualitative content, making it unlikely that judges were exposed to different kinds of cues. However, it may be possible that judges' inferences were shaped by the specific context in which they were interacting, rather than solely by the range of cues that were available to them. Future research could clarify this possibility by employing a 'think-aloud' protocol to derive an understanding of how judges extract relevant cues from linguistic and nonverbal content.

A second factor worth acknowledging concerns the method used to examine judgment 'accuracy' and the size of the resulting correlations. The correlations found in this study, which ranged from .01 to .20, are not atypical of research where judgments are made across judge-target dyads. For example, Back et al.'s (2008) analysis of Big-5 personality judgments made from email address found correlations ranging from .05 to .13, while Naumann et al.'s (2009) analysis of Big-5 judgments made from physical appearance cues found correlations ranging from -.01 to .29. However, in absolute terms, they are low, and it is difficult to determine the extent to which this reflects a negligible finding or the consequence of removing substantive variance while seeking to control for the mathematical artifacts identified by Cronbach (1955).<sup>5</sup> One approach to mitigating this concern is to demonstrate that judge-target correlations of this magnitude do allow for good prediction of whether or not a target will undertake a subsequent behavior (e.g., Oshin, Stewart, Mandel, Pavlovic, McLellan, & Taylor, 2012). In future work it should be possible to demonstrate that relative accuracy in judgments across contexts translates to relative accuracy at predicting an external behavior by the target.

The present study has provided the first systematic exploration of the impact of context richness on the accuracy of personality judgments using Big-5 accuracy scores. The findings have revealed that extraversion and neuroticism are judged more accurately in 'rich' interaction

contexts where there is a breadth of verbal, paralinguistic, and nonverbal cues, whereas conscientiousness and openness are judged more accurately in information-lean contexts. This suggests that there may be an optimal fit between ‘everyday’ contexts and the trait being judged. In a world where our judgments of others are increasingly made on the basis of information produced through an expanding range of mediums—some lean, some rich, some single-channel, some multi-channel, and so on—the idea that rich contexts are not necessarily better is of fundamental significance to our understanding of interpersonal perception.

### References

- Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs*, *35*, 492-503. doi:10.1080/03637756809375599
- 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
- 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*(3), 431-441. doi: 10.1037/0022-3514.64.3.431
- Ashton, M. C., & Lee, K. (2005). Honesty-humility, the big five, and the five factor model. *Journal of Personality*, *73*, 1321-1353. doi: 10.1111/j.1467-6494.2005.00351
- Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality traits from email addresses. *Journal of Research in Personality*, *42*, 1116-1122. doi: 10.1016/j.jrp.2008.02.001
- Bagby, R. M., Rector, N. A., Bindseil, K., Dickens, S. E., Levintan, R. D., & Kennedy, S. H. (1998). Self-report ratings and informants' ratings of personalities of depressed outpatients. *The American Journal of Psychiatry*, *155*, 437-438.
- Blackman, M. C., & Funder, D. C. (2002). Effective interview practices for accurately assessing counterproductive traits. *International Journal of Selection and Assessment*, *10*, 109-116. doi:10.1111/1468-2389.00197
- 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

- Buchanan, T., Johnson, J. A., & Goldberg, L. (2005). Implementing a Five-Factor personality inventory for use on the Internet. *European Journal of Psychological Assessment, 21*, 115-127. doi: 10.1027/1015-5759.21.2.115
- 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: org/10.1016/j.jrp.2007.01.004
- Christiansen, N. D., Wolcott-Burnam, S., Janovics, G. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance, 18*, 123-149. doi: 10.1207/s15327043hup1802\_2
- Cronbach, L. J. (1955). Processes affecting scores on “understanding of others” and “assumed similarity”. *Psychological Bulletin, 52*, 177-193. doi: 10.1037/h0044919
- Culnan, M. J., & Markus, M. L. (1987). Information technologies. In F. M. Jablin, L. L. Putnam, K. H. Roberts, & L. W. Porter (Eds.), *Handbook of organizational communication: An interdisciplinary perspective* (pp. 420-443). Newbury Park, CA: Sage Publications.
- Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*, 554-571. doi: 10.1287/mnsc.32.5.554
- Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. T. (2011). Mark my words!: linguistic style accommodation in social media. Proceedings of the 20th international conference on World wide web, 745-754. doi:10.1145/1963405.1963509
- Funder, D. C. (1999). *Personality judgment: a realistic approach to person perception*. San Diego, CA: Academic Press.

- Funder, D. C., & Colvin, 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., & Colvin, R. (1997). Congruence of self and others' judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617-647). Orlando, FL: Academic Press.
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with inter-judge agreement. *Journal of Personality and Social Psychology*, *52*, 409-418. doi: 10.1037/0022-3514.52.2.409
- Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgment accuracy. *Journal of Personality and Social Psychology*, *64*, 479-490. doi: 10.1037/0022-3514.64.3.479
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96. doi:10.1016/j.jrp.2005.08.007
- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A Room with a cue: Judgments of personality based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*, 379-398. doi: 10.1037/0022-3514.82.3.379
- Gosling, S. D., Gaddis, S., & Vazire, S. (2007, March). *Personality impressions based on Facebook profiles*. Paper presented at the International Conference on Weblogs and Social Media, Boulder, CO. Downloaded on 29 July 2011, from: <http://www.icwsm.org/papers/3--Gosling-Gaddis-Vazire.pdf>

- Hancock, J. T., & Dunham, P. J. (2001). Impression formation in computer mediated communication revisited. *Communication Research, 28*, 325-347. doi: 10.1177/009365001028003004
- Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analysis. *Psychological Bulletin, 105*, 302-308. doi: 10.1037/0033-2909.105.2.302
- International Personality Item Pool: A scientific collaboratory for the development of advanced measures of personality traits and other individual differences* (2011). Retrieved October 31, from <http://ipip.ori.org/>
- Kenny, D. A. (1995). The effect of non-independence on significance testing in dyadic research. *Personal Relationships, 2*, 67-75. doi: 10.1111/j.1475-6811.1995.tb00078.
- Kiesler, S. (1986). The hidden messages in computer networks. *Harvard Business Review, issue 1*, 46-58.
- Kurtz, J. E., & Putnam, S. H. (2006). Patient-informant agreement on personality ratings and self-awareness after head injury. *Clinical Neuropsychology, 20*, 453-468. doi: 10.1080/13854040590967090
- 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
- Li, J., & Chignell, M. H. (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human Compute Studies, 68*, 589-602. doi: 10.1016/j.ijhcs.2010.04.001

- Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Quantity and quality of available information affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91*, 111-123. doi: 10.1037/0022-3514.91.1.111
- Markey, P. M., & Wells, S. M. (2002). Interpersonal perception in Internet chat rooms. *Journal of Research in Personality, 36*, 134-146. doi: org/10.1006/jrpe.2002.2340
- Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology, 42*, 250-261.
- 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
- Vartanian, O., Stewart, K., Mandel, D. R., Pavlovic, N., McLellan, L., & Taylor, P. J. (2012). Personality assessment and behavioral prediction at first impression. *Personality and Individual Differences, 52*, 250-254. doi: org/10.1016/j.paid.2011.05.024
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count (LIWC2007): A computer-based text analysis program [computer software]. Austin, TX: LIWC, Inc.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*, 1296-1312. doi: 10.1037/0022-3514.77.6.1296



- Scherer, K. R., Scherer, U., Hall, J. A., & Rosenthal, R. (1977). Differential attribution of personality based on multi-channel presentation of verbal and nonverbal cues. *Psychological Research, 39*, 221-247. doi:10.1007/BF00309288
- Shevlin, M., Walker, S., Davies, M. N. O., Banyard, P., & Lewis, C. A. (2003). Stranger ratings of personality from photographs: Consensus and accuracy. *Personality and Individual Differences, 35*, 1373-1383. doi:10.1016/S0191-8869(02)00356-2
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428. doi: 10.1037/0033-2909.86.2.420
- Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech, 18*, 145-152.
- 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
- Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*, 221-229. doi: 10.1016/j.jrp.2006.03.001
- Wall, H. J. (2011). Conversation with a Cue: Personality Judgements and Observer Accuracy across Context. Unpublished Doctoral Thesis.
- Walther, J. B., Slovacek, C., & Tidwell, L. C. (2001). Is a picture worth a thousand words? Photographic images in long term and short term virtual teams. *Communication Research, 28*, 105-134. doi: 10.1177/009365001028001004
- Watson, D. (2000). *Mood and temperament*. New York: Guilford Press.

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44, 363-373. doi:

10.1016/j.jrp.2010.04.001

Yee, N., Harris, H., Jabon, M., Bailenson, J.N. (2011). The expression of personality in virtual worlds. *Social Psychology and Personality Science*, 2 (1), 5-12. doi.

10.1177/1948550610379056

### Footnotes

- <sup>1</sup> Interestingly, in Funder and Sneed's study, agreeableness was associated with both content (e.g., acts irritably) and visual behaviors (e.g., behaviors in a cheerful manner). For this reason we did not consider this trait any further.
- <sup>2</sup> Although pairs of participants were used, some final participant numbers are odd due to missing data. As the unit of analysis was the individual not the dyad, odd numbers do not pose a problem to the findings.
- <sup>3</sup> We included agreeableness items within the IPIP measure for completeness, but since our hypotheses concern trait relevance and visibility, we made the a priori decision to exclude agreeableness from our analyses (see also footnote 1). The accuracy of judgments for this trait showed a small, non-significant linear decrease as context richness increased from Internet 'chat' ( $r = .08$ ), to telephone ( $r = .05$ ), to face-to-face ( $r = .03$ ),  $F < 1$ , *ns*.
- <sup>4</sup> The term accuracy is not put forward as a 'gold standard' in assessing the validity of personality judgements. Rather, it is used for intuitive appeal and for consistency with previous literature on the topic of realistic personality assessment and accuracy (e.g., Funder, 1999). Accuracy as opposed to realistic accuracy is used here as the present research uses an aggregate of self and informant ratings whereas realistic accuracy uses an average of self, informant, and clinicians' ratings (Funder, 1999).
- <sup>5</sup> Analyses examining the unstandardized accuracy scores (i.e., before our standardization was applied) found comparable results but with larger coefficients (e.g., the coefficients for extraversion were .02, .18, and .30 for Internet 'chat', telephone, and face-to-face, respectively). Across the traits, the range of coefficients observed for the unstandardized scores was .01 – .30, while the range of our standardized coefficients was -.05 – .20.

Figure 1. Interactant judgment accuracy as a function of interaction context and personality trait.

