Rating E-mail Personality at Zero Acquaintance*

Alastair Gill[a]
Jon Oberlander[a]
Elizabeth Austin[b]

a School of Informatics, University of Edinburgh, 2 Buccleuch Place, Edinburgh, EH8 9LW, UK

b School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

* We are grateful to all those who have commented on this work, and to the ESRC and the University of Edinburgh for their support. Queries should be addressed to Alastair Gill or Jon Oberlander. Email addresses: A.Gill@ed.ac.uk, J.Oberlander@ed.ac.uk , Elizabeth.Austin@ed.ac.uk.

Wordcount: 4,997

Abstract

Electronic media are pervasive in daily communication. But how well can personality traits be perceived through a short e-mail message? Working independently and under experimenter supervision, thirty judges each rated 18 short e-mail texts. These texts were written under experimental conditions by authors of known personality, who briefly described their recent activities, and were collected as part of a previous study, which demonstrated linguistic projection of personality. As predicted by the literature, we find that even with minimal textual cues there is relatively high agreement for ratings of author Extraversion, and lesser agreement for Psychoticism. However, agreement for Neuroticism ratings, especially between target and judges, appears to be further reduced by the environment. We suggest that these judgements may be mediated by stereotypes. In addition to reducing the cues available for personality rating, the study extends previous work in two main ways: it uses the three-factor model of personality, and thus measures Psychoticism; and it adopts exemplar-based and subjective measures of personality perception.

Key words: Personality perception, Language, Computer-mediated communication, Subjective measures

1 Introduction

When we express ourselves through language, how much can others tell about our personality? What happens when we express ourselves via computer-mediated communication (CMC), and e-mail in particular? Although e-mail is a written form of language, it has many similarities with spoken interaction (Colley and Todd, 2002). Given its popularity, understanding how personality is projected and perceived through e-mail is a timely issue (Baron, 1998).

In face-to-face interaction we are highly effective at judging personality (*e.g.*, Funder and Dobroth, 1987; Funder and Colvin, 1988; Paunonen, 1989). However, e-mail is frequently used to make first contact with people, while lacking many of the cues usually used for face-to-face personality judgement. Synchronic—*i.e.* simultaneous—CMC is already known to have implications for personality judgement (Hancock and Dunham, 2001; Markey and Wells, 2002), and so here we study the effects of asynchronous e-mail.

We use Eysenck's three-factor model of personality (EPQ-R; Eysenck and Eysenck, 1991) with its dimensions of Extraversion, Neuroticism and Psychoticism; however, where relevant, we discuss connections to five-factor models (Digman, 1990; Costa and McCrae, 1992), and its dimensions of Extraversion, Neuroticism, Conscientiousness, Agreeableness and Openness.

2 Background

2.1 Perception of Personality

Personality judgements can be derived from self-report, or the perceptions of family, friends or strangers. Here we discuss relevant literature in relation to Funder's Realistic Accuracy Model, which grounds the processes of accurate personality judgement in terms of the quality of the 'judge', 'target', 'trait', and 'information' (cf. Kenny, 1994; Funder, 1995).

Good judges are distinguished from bad by their more effective use of the cues which are available to them. So better socialised judges, such as Extraverts, will be better judges (Akert and Panter, 1988).

Good targets are those whose behaviour makes available numerous and informative clues to their personality. Targets with higher levels of social behaviours exhibit more potential clues about their personality (*e.g.* Borkenau and Liebler, 1992).

Good traits are highly visible, and demonstrate low evaluativeness. Funder notes that Extraversion is highly visible and revealed by 'frequent positive social interaction' (Funder and Dobroth, 1987; Funder and Colvin, 1988; Paunonen, 1989), but relatively low in evaluativeness, or affective charge. However, Neuroticism is lower in visibility (*e.g.* worrying thoughts or feelings), and more evaluative. This may lead to: the concealment of undesirable behaviour from observers; a distortion of self-perception, leading to lower target-judge agreement; or a greater reluctance to judge such behaviours, leading to reduced inter-judge agreement. When less evaluative measures of Neuroticism are used, agreement increases (John and Robbins, 1993).

Certainly, even in judgements by close acquaintances, much greater agreement is found for ratings of Extraversion than for Neuroticism in both the EPQ, and in five factor models. For the EPQ, Psychoticism has shown marginally lower agreement than Neuroticism (Goma-i-Freixanet, 1997). For the five-factor model, Openness generally shows similar levels of agreement to Extraversion, whereas Agreeableness shows low agreement similar to that of Neuroticism; Conscientiousness is located somewhere between these groups (*e.g.* McCrae and Costa, 1987).

Good information about targets depends upon both quantity and relevance, and influences the judges' levels of agreement. Close acquaintances agree better with each other and with the target, than do other peers, or relative strangers (Funder and Colvin, 1988; Paunonen, 1989; Paulhus and Bruce, 1992). At zero-acquaintance, strangers rating personality using minimal cues (linguistic or visual) show greater agreement for more visible traits such as Extraversion or Conscientiousness (Gifford and Hine, 1994; Albright *et al.*, 1988). However, given the scarcity of information, these judgements are often influenced by stereotypes relating to attractiveness, nationality, gender, or even personality (Scherer, 1972; Gallois and Callan, 1986; McCrae and Costa, 1987).

2.2 Computer-mediated communication

2.2.1 Effects of CMC on language

CMC, and more specifically e-mail, is considered to be a form of communication located between the domains of speech and writing (Baron, 1998): It is written, with interlocutors physically separated; it is durable, and often utilises complex syntactic constructions. However, e-mail is often unedited, makes extensive use of first and second person pronouns, present tense and contractions, and is informal. Colley and Todd (2002) note specific stylistic 'emailisms', including ellipses, capitalisation, excessive use of exclamation marks and question marks. Multi-dimensional linguistic analysis of a bulletin board corpus found that the genre resembled that of 'interviews and letters' (Collot and Belmore, 1996).

Although CMC is impoverished compared to face-to-face interaction, it can encourage increased communication from shy individuals (Bloch, 2002), and also provide relatively rich information about communicators, such as their status or gender (*e.g.* Panteli, 2002; Colley and Todd, 2002). Outside of CMC, target personality and general language use are known to be linked (*e.g.* Scherer, 1979; Furnham, 1990; Pennebaker and King, 1999): personality influences both syntax (Dewaele, 2001), and lexical choice (Pennebaker and King, 1999). Within CMC, corpus-based work has confirmed the persistence of such patterns, and revealed links between personality and emailisms (Gill and Oberlander, 2002).

2.2.2 Effects of CMC on personality judgement

When the amount of information for personality judgements is reduced, accuracy falls too. Zero-acquaintance judgements are particularly vulnerable to technological artifacts. For example, interviews conducted by telephone result in reduced self-interviewer and peer-interviewer agreement than face-to-face interviews (Blackman, 2002). In text-based CMC, accuracy in judgements of gender was reduced by expectations of linguistic stereotypes for the male and female writers (Savicki *et al.*, 1999). For judgements of personality following one-on-one interaction in an internet chat room, judges agreed over a target's Extraversion, Agreeableness, and Openness, but target-judge agreement was only found for Extraversion and Openness

(Markey and Wells, 2002). Impressions of personality following task-oriented synchronous CMC have been found to be less detailed but more intense compared with those from face-to-face communication. Judges seemed less able to rate their partners for Extraversion, Neuroticism, and Agreeableness (Hancock and Dunham, 2001).

2.3 Perception Hypotheses

Given this discussion, we set out to test these hypotheses:

Psychoticism will show reduced agreement, due to its high evaluativeness, low visibility, and the reduced information available in the CMC at zero-acquaintance.

Extraversion will be the most easily perceived due to its high visibility and low evaluativeness, and so will show the highest levels of agreement, even in CMC at zero-acquaintance.

Neuroticism will show agreement similar to Psychoticism, for the same reasons.

3 Method

3.1 Judges

The 30 judges were undergraduate or postgraduate students, or recent graduates currently living in Edinburgh (15 males, 15 females; mean age=21.6 years, SD=1.24). All were experienced e-mail users (on a scale of 1–10, with 10 being ‘a great deal’; mean=9.23, SD=0.77), and all were naive raters of personality (18 had no experience of psychology; 12 had ‘some’ experience—having read books on it, or studied it as part of their degree). None had previously taken part in any personality rating experiments.

3.2 Materials

The rating booklet sections were similarly structured for each personality trait: First a description of the personality trait was given, and then on each subsequent page after an introduction to the task, there was a target text followed by several questions relating to the judge's perception of the text's author.

The target texts were taken from an e-mail corpus collected previously (Gill, 2003). This consisted of 210 texts, 2 each from 105 subjects, who had completed the EPQ-R short form (Eysenck and Eysenck, 1991). The texts had been written under experimental conditions, as if to a good friend. From this set, 18 were selected for the experiment; 6 texts were chosen to represent a range of author personality for each dimension. Extreme high and low personality scores were defined as those greater than 1 standard deviation of the mean, and two texts represented each of these. Two further texts were selected—one above and below the mean—to represent less extreme realisations of the trait (each between .5 and 1 SD of the mean). In each case, the scores for the other personality dimensions were controlled for, being $<\pm 1$ SD of the mean (in most cases $<\pm .5$ SD). Mean lengths in words were: Psychoticism texts=258.67, Extraversion texts=261.33, Neuroticism texts=261.00.

The rating questionnaire was divided into three sections, one for each personality trait (Psychoticism, Extraversion, or Neuroticism); the terms Tough-mindedness and Emotionality were used instead of Psychoticism and Neuroticism. The texts for each dimension were arranged in random order, and the order for each dimension determined by a Latin square technique. Booklets were then given an identification code, to maintain judge anonymity.

The booklet was prefixed by an introduction emphasising our interest in ‘how subjects think the author comes across’, the need for them to answer ‘honestly and accurately’ and ‘not to spend too long thinking about each question’. For each personality dimension a description based upon those of Eysenck and Eysenck (1991) was used, with minor re-wording to enhance intelligibility and minimise issues of social desirability. Although it is more usual to rate personality using a standard set of questions (Eysenck and Eysenck, 1991; Costa and McCrae, 1992; cf. Ten-Item

Personality Inventory, Gosling *et al.*, 2003), Sneed et al. (1998) have found that factors and their manifestation through behaviour can easily be identified using exemplars.

Judges were first asked 'How [Tough-Minded/Extravert/Emotionally-Stable] is the author of the e-mail?', with the extremes of the scale labelled 'Not at All' and 'Very [Tough-Minded/Extravert/Emotionally-Stable]'. They were then asked 'How similar would you say is the personality of this e-mail's author to yours?' rated on a scale of 1–10, labelled 'Very Different' and 'Very Similar' respectively.

3.3 Procedure

Judges worked through the rating booklet at their own speed. Although there was no official time limit, they were encouraged to work 'quickly and efficiently' to minimise reflection and retain motivation. Several judges participated in the experiment at the same time, under exam-type conditions, over-seen by the experimenter. Equal numbers of participants were randomly assigned to each questionnaire. After completing the rating booklet, there was a short debriefing session which included administration of the EPQ-R (Eysenck and Eysenck, 1991).

4 Results

4.1 Judges

The judges' completion of the EPQ-R (short form; Eysenck and Eysenck, 1991) gave the following results: Psychoticism Mean score: 3.17, SD 2.4; Extraversion Mean score: 7.30, SD 2.6; Neuroticism Mean score: 5.30, SD 3.1; and Lie Scale Mean score: 3.27, SD 2.0. These indicate that the judges' personality profile is similar to the published norms.

4.2 Consistency and Agreement of Judges' Ratings

All 6 authors for each of the three personality traits were scored on a scale of 1–10 by each judge. Concordance between the judges was measured using Kendall's *W*, and in all cases the

Kendall coefficient reached a level of statistical significance, indicating relative agreement among judges concerning the trait score of each text (Psychoticism, 0.287 [$W(5)=43.05$; $p<0.0001$]; Extraversion, 0.471 [$W(5)=70.64$; $p<0.0001$]; Neuroticism, 0.266 [$W(5)=38.91$; $p<0.0001$]). Additionally we examine how each judge agrees with each of the others (cf. Morris *et al.*, 2002). Therefore, correlations were performed for each judge with each of the other judges, and the mean correlation reported for each judge (counts of correlations achieving significance are also noted for each cell out of a maximum of 29). Given the ordinal nature of the rating scale responses, Spearman rank correlations are used throughout these analyses.

The final row of Table 1 gives the average rank correlation for each trait across all judges. Extraversion is shown to have the greatest inter-judge agreement, and in these terms appears to be the easiest trait to rate (mean r_s=0.482). Psychoticism (mean r_s=0.333), and Neuroticism (mean r_s=0.308) both show lower levels of agreement, which suggests that they are harder to rate. The greater agreement for Extraversion is also reflected in the total number of significant correlations found for the trait (76), which exceeds the totals for either Psychoticism (26) or Neuroticism (20).

Insert Table-1

Since we calculate Spearman rank correlations, we have reported the means of these correlations (Morris *et al.*, 2002), rather than using Fischer's *r* to *z* conversion (*e.g.*, Funder and Colvin, 1988; Vogt and Colvin, 2003). To establish the significance of agreement between judges, intraclass correlations were calculated across the thirty judges for their ratings of P, E, and N targets (the equivalent of performing correlations between all possible pairs of raters; McCrae and Costa, 1987). Again, Extraversion showed the highest agreement, with an intraclass correlation of 0.403, with relatively lower agreement for Psychoticism (0.206), and Neuroticism (0.248); all significant at $p<0.0001$.

4.3 Target-Judge Correlation

To see how the individual judges had performed, mean correlations of judge-target agreement were calculated. For each of the judges, their text ratings were correlated with the original personality scores of the authors, and their mean performance for rating P, E, and N was also noted (Table 2). The largest number of significant correlations (out of a possible 30) were found for Extraversion (5), followed by Psychoticism (2); none reached significance for Neuroticism.

We calculate the aggregate measure of personality ratings across multiple raters since McCrae and Costa (1987) suggest that this takes into account how the target is seen by the judgement group overall. Thus, we correlated the targets' raw EPQ-R score and mean judges' ratings (1–10): Extraversion r_s=0.886; Psychoticism r_s=0.754; Neuroticism r_s=–0.377 (Spearman, pairwise, two-tailed, 6 cases). Of these, only ratings of Extraversion showed significant target-judge agreement (p<0.05; however Psychoticism demonstrated significance at the p<0.1 level).

Insert Table-2

4.4 Quality of Judges and Targets

The level of agreement between judges across all three personality traits is also shown in Table 1. Here, judges 20 and 30 agree most with the other judges, and judges 5 and 19 the least; the mean level of agreement across P, E, and N dimensions was 0.374. Agreement between target and judge ratings across all dimensions can be found in Table 2. Here judges 21 and 17 show the greatest agreement, with judges 8 and 12 actually showing negative agreement with the targets.

Turning to the target texts, if one text for a personality trait is particularly difficult to rate, then we expect greater variability in judges' ratings for it. Levene's test for homogeneity of variance was used to investigate this. Although significant differences were not found for Extraversion or Neuroticism, they were found for Psychoticism [F(5,174)=2.868, p<0.05]. High Psychotic texts

(2 high-P, and the mid-high-P) consistently showed the greatest variance in ratings, which suggests that the judges found these the most difficult to rate.

4.5 Perceived Similarity of Targets to Judges

Analysis of the similarity ratings allows us to investigate how judges perceived the target author personalities relative to their own.

The six target texts for each personality dimension were grouped into categories of High, Mid, and Low. A within subjects analysis of variance (ANOVA) revealed effects of text personality type on ratings of similarity for Psychoticism [$F(2,58)=7.999$, $p<0.001$, $MSE=1.6$], and for Extraversion [$F(2,58)=4.052$, $p<0.05$, $MSE=1.6$], but not Neuroticism texts. Tukey HSD tests revealed that significant differences in similarity ratings were found between LowP (M=5.6) and HighP (M=4.3), and also HighP (M=4.3) and MidP (M=5.1) Psychoticism texts, and between the HighE (M=5.3) and MidE (M=4.3) Extraversion texts (all significant at $p<0.05$).

To take into account the effects of judge personality on the ratings of similarity, judges were then categorised as either ‘high’ or ‘low’ on the personality dimension in question using a mean split. A two factor mixed-design ANOVA revealed for Psychoticism main effects of judge personality type [$F(1,28)=6.555$, $p<0.05$, $MSE=3.1$] and (as would be expected) personality of text author [$F(2,56)=8.063$, $p<0.001$, $MSE=1.6$]. However, no interaction effect was found between judge personality and text author personality in the ratings of similarity. Within subjects ANOVA of only the low Psychotic judges shows an effect of text type on similarity rating [$F(2,32)=5.753$, $p<0.01$, $MSE=1.9$]. Tukey tests revealed significant results ($p<0.05$), with low Psychotic judges rating themselves as most similar to the LowP texts (M=6.2), and most dissimilar to the HighP texts (M=4.6). The lack of interaction noted above can be attributed to high Psychotic judges' tendency also to rate themselves as most dissimilar to the HighP texts; but this effect was only significant at $p<0.1$.

The two factor mixed-design ANOVA revealed for Extraversion a main effect for text personality type on similarity rating [$F(2,56)=4.390$, $p<0.05$, $MSE=1.5$], and also an interaction

effect for rater and text personality upon similarity ratings [$F(2,56)=3.430$, $p<0.05$, $MSE=1.5$]. Within subjects ANOVA of only the high Extravert judges shows (as expected from the significant interaction) effects of text type on their similarity ratings [$F(2,26)=5.082$, $p<0.05$, $MSE=1.9$]. Tukey tests reveal significant effects ($p<0.05$): the high Extravert judges rated the HighE texts as most similar to themselves (M=6.1) and the MidE texts as least similar (M=4.5). No effects were found for Neuroticism.

5 Discussion

The expectation was that the visibility and evaluativeness of personality traits would influence how they are perceived via CMC at zero-acquaintance. In particular, it was expected that Extraversion would be the easiest to perceive, with Psychoticism and Neuroticism more difficult.

Extraversion shows the greatest inter-judge agreement, and the best target-judge agreement, with a strong and significant correlation. This is consistent with previous findings, and reflects Extraversion's more observable and less evaluative properties. Neuroticism and Psychoticism both show lower levels of inter-judge agreement. However, turning to target-judge agreement, although both levels of agreement are lower than for Extraversion, Psychoticism actually shows a strong positive correlation approaching significance whilst Neuroticism demonstrates a weak and non-significant *negative* correlation.

The results differ from both John and Robbins (1993)'s five factor study, where Neuroticism shows quite good agreement, compared with Conscientiousness and Agreeableness, and from Goma-i-Freixanet (1997)'s three factor study, where Neuroticism had marginally better target-judge agreement than Psychoticism. We can only compare directly with the latter study. In Funder's Realistic Accuracy terms, it is clear that Neuroticism is a ‘bad’ trait, with higher evaluativeness and lower visibility compared with Extraversion. But how do our Neuroticism results relate to those for Psychoticism?

The perceived similarity results bear on this question. For Extraversion, the high target-judge agreement already suggests the trait is visible and accurately perceived. The interaction effect in

the similarity judgements demonstrate that the trait is less evaluative, despite the apparent desirability which leads the judges as a whole to rate themselves as more similar to the high Extravert targets. By contrast, target-judge agreement is still high for Psychoticism. Given its visibility, the lack of an interaction effect in similarity judgements would result from this trait's greater evaluativeness. Now consider the low target-judge agreement for Neuroticism. If the trait is evaluative, then the lack of any discernible pattern in the similarity judgements suggests that low visibility is an interfering factor. Perhaps e-mail fails to provide 'good information' for accurate judgement of this less visible trait.

But previous studies of the e-mail data have shown that linguistic features of Neuroticism are projected in CMC (Gill, 2003). So we suggest that the judges are attending to the wrong information. CMC judgements are known to be susceptible to stereotype effects (Savicki *et al.*, 1999). In a speech study, Scherer (1972) found that despite the high level of inter-rater reliability for Extraversion, there was little target-judge agreement. He concluded that judges were instead attending to stereotyped cue information, influenced by social desirability. In our CMC case, for Neuroticism, judges may be attending to misinformed stereotyped cues, perhaps primed by the exemplars used for rating. Certainly, these descriptions of Neuroticism may increase the evaluativeness of the measure for this trait (cf. John and Robbins, 1993). But the use of exemplars is not in itself problematic, since judgements of Extraversion and Psychoticism are not apparently mediated by stereotypes. So people may have access to a better-formed stereotype of what constitutes Neurotic behaviour, and at zero-acquaintance, attend to this, instead of 'good information'.

6 Conclusion

From a text of around 300 words, 30 judges were able to consistently agree (both with each other and with the target individual's self-rating), on the personality of the text's author when rating them for Extraversion and to a slightly lesser extent, for Psychoticism. In both cases, judges used an exemplar-based rating of personality rather than an itemised personality questionnaire. In addition, judges rated perceived target similarity. Although judges generally agreed with each other regarding ratings of Neuroticism, unexpectedly little consistency was found with the

targets' own personality assessments, or with perceived similarity. We propose that this is due to characteristics of the trait itself, interacting with the CMC environment at zero-acquaintance, which leads judges to attend to the wrong, possibly stereotyped, information.

References

Akert, R. and Panter, A. (1988). Extraversion and the ability to decode nonverbal communication. *Personality and Individual Differences*, **9**, 965–972.

Albright, L., Kenny, D., and Malloy, T. (1988). Consensus in personality judgements at zero acquaintance. *Journal of Personality and Social Psychology*, **55**(3), 387–395.

Baron, N. (1998). Letters by phone or speech by other means: the linguistics of email. *Language and Communication*, **18**, 133–170.

Baron, N. (2001). Commas and canaries: the role of punctuation in speech and writing. *Language Sciences*, **23**, 15–67.

Blackman, M. (2002). The employment interview via the telephone: Are we sacrificing accurate personality judgements for cost efficiency. *Journal of Research in Personality*, **36**, 208–223.

Bloch, J. (2002). Student/teacher interaction via email: the social context of internet discourse. *Journal of Second Language Writing*, **11**, 117–134.

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

Colley, A. and Todd, Z. (2002). Gender-linked differences in the style and content of e-mails to friends. *Journal of Language and Social Psychology*, **21**, 380–392.

Collot, M. and Belmore, N. (1996). Electronic language: A new variety of English. In S. Herring, (Ed), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, pp. 13–28. Amsterdam: Benjamins.

Costa, P. and McCrae, R.R. (1992). *NEO PI-R Professional Manual*. Odessa, Florida: Psychological Assessment Resources.

Dewaele, J.-M. (2001). Interpreting the maxim of quantity: interindividual and situational variation in discourse styles of non-native speakers. In E. Németh, (Ed), *Selected Papers from the 7th International Pragmatics Conference, volume 1*, pp. 85–99. Antwerp: International Pragmatics Association.

Digman, J. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, **41**, 417–440.

Eysenck, H. and Eysenck, S.B.G. (1991). *The Eysenck Personality Questionnaire-Revised.* Sevenoaks: Hodder and Stoughton.

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

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

Funder, D.C. (1995). On the accuracy of personality judgement: A realistic approach. *Psychological Review*, **102**, 652–670.

Furnham, A. (1990). Language and personality. In H. Giles and W. Robinson, (Eds), *Handbook of Language and Social Psychology*, pp. 73–95. Chichester; Wiley.

Gallois, C. and Callan, V. (1986). Decoding emotional messages: Influences of ethnicity, sex, message type, and channel. *Journal of Personality and Social Psychology*, **51**, 755–762.

Gifford, R. and Hine, D.W. (1994). The role of verbal behaviour in the encoding and decoding of interpersonal dispositions. *Journal of Research in Personality*, **28**, 115–132.

Gill, A. (2003). *Personality and Language: The projection and perception of personality in computer-mediated communication*. PhD thesis, University of Edinburgh, Edinburgh, UK.

Gill, A. and Oberlander, J. (2002). Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pp. 363–368. Fairfax, VA, August 2002.

Goldberg, L. (1993). The structure of phenotypic personality traits. *American Psychologist*, **48**, 26–34.

Goma-i-Freixanet, M. (1997). Consensus validity of the EPQ: Self-reports and spouse-reports. *European Journal of Psychological Assessment*, **13**(3), 179–185.

Gosling, S.D., Rentfrow, P.J., and Swann, W.B. (2003). A very brief measure of the big five personality domains. *Journal of Research in Personality*, **37**, 504–528.

Hancock, J. and Dunham, P. (2001). Impression formation in computer-mediated communication. *Communication Research*, **28**, 325–347.

John, O. and Robbins, R. (1993). Determinants of interjudge agreement: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, **61**, 521–551.

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

Markey, P. and Wells, S. (2002). Interpersonal perception in internet chat rooms. *Journal of Research in Personality*, **36**, 134–146.

McCrae, R. and Costa, P. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, **52**, 81–90.

Morris, P., Gale, A., and Duffy, K. (2002). Can judges agree on the personality of horses? *Personality and Individual Differences*, **33**, 67–81.

Panteli, N. (2002). Richness, power cues and email text. *Information and Management*, **40**, 75–86.

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

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

Pennebaker, J.W. and King, L. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, **77**, 1296–1312.

Savicki, V., Kelley, M., and Oesterreich, E. (1999). Judgements of gender in computer-mediated communication. *Computers in Human Behavior*, **15**, 185–194.

Scherer, K. (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, **40**, 191–210.

Scherer, K. (1979). Personality markers in speech. In K.R. Scherer and H. Giles, (Eds), *Social Markers in Speech*, pp. 147–209. Cambridge: Cambridge University Press.

Sneed, C., McCrae, R., and Funder, D. (1998). Lay conceptions of the Five-Factor Model and its indicators. *Personality and Social Psychology Bulletin*, **24**, 115–126.

Vogt, D. and Colvin, R. (2003). Interpersonal orientation and the accuracy of personality judgements. *Journal of Personality*, **71**, 267–295.

Judge	Psychoticism		Extraversion		Neuroticism		Mean r_s
1	0.396	(2)	0.199	(1)	–0.007	(0)	0.196
2	0.227	(0)	0.407	(0)	0.448	(1)	0.361
3	0.176	(0)	0.497	(2)	0.351	(1)	0.341
4	0.489	(0)	0.367	(0)	0.230	(0)	0.362
5	–0.142	(0)	0.014	(0)	0.466	(0)	0.113
6	0.482	(2)	0.594	(5)	0.253	(1)	0.443
7	0.378	(1)	0.682	(6)	0.341	(1)	0.467
8	0.362	(0)	0.155	(1)	0.090	(0)	0.202
9	0.413	(2)	0.533	(3)	0.246	(1)	0.397
10	0.309	(0)	0.537	(3)	0.442	(1)	0.429
11	0.367	(0)	0.666	(4)	0.220	(0)	0.418
12	0.333	(1)	0.422	(0)	0.300	(1)	0.352
13	0.092	(0)	0.429	(0)	0.490	(2)	0.337
14	0.493	(0)	0.178	(0)	0.540	(0)	0.404
15	0.510	(2)	0.400	(0)	0.237	(1)	0.382
16	0.463	(2)	0.314	(0)	0.285	(1)	0.354
17	0.380	(0)	0.501	(2)	0.383	(1)	0.421
18	0.327	(1)	0.520	(2)	0.299	(1)	0.382
19	0.100	(0)	0.569	(1)	–0.086	(0)	0.194
20	0.379	(2)	0.652	(6)	0.531	(1)	0.521
21	0.369	(1)	0.562	(2)	0.267	(0)	0.399
22	0.218	(1)	0.581	(6)	0.459	(0)	0.419
23	0.298	(1)	0.320	(0)	0.436	(1)	0.351
24	0.176	(0)	0.682	(7)	0.417	(1)	0.425
25	0.288	(0)	0.626	(7)	0.352	(1)	0.422
26	0.471	(3)	0.666	(6)	0.175	(1)	0.437
27	0.340	(1)	0.642	(3)	–0.112	(0)	0.290
28	0.403	(1)	0.541	(2)	0.449	(0)	0.464
29	0.429	(0)	0.602	(2)	0.349	(0)	0.460
30	0.472	(3)	0.613	(5)	0.374	(2)	0.486
Mean r_s	0.333		0.482		0.308		0.374

Note: Agreement is described by the mean correlation of each judge with other judges for each scale. The number of statistically significant positive correlations (at the $p < 0.05$ level) is shown in brackets, maximum 29 per cell.

Table 1: Inter-Judge Agreement correlations for raters

Judge	Psychoticism	Extraversion	Neuroticism	Mean r_s
1	0.729	–0.114	–0.186	0.143
2	0.200	0.714	–0.614	0.100
3	–0.200	0.700	0.386	0.295
4	0.571	0.314	–0.257	0.209
5	–0.229	0.329	0.100	0.067
6	0.771	0.829	0.157	0.586
7	0.386	0.886	0.300	0.524
8	0.071	–0.143	–0.329	–0.134
9	0.586	0.714	0.214	0.505
10	0.000	0.814	–0.243	0.190
11	0.500	0.800	0.429	0.576
12	0.114	0.286	–0.557	–0.052
13	0.171	0.329	–0.486	0.005
14	0.929 *	0.329	0.343	0.534
15	0.686	0.629	–0.229	0.362
16	0.543	0.457	0.157	0.386
17	0.829	0.757	0.300	0.629
18	0.357	0.814	0.500	0.557
19	0.214	0.700	0.443	0.452
20	0.629	0.986 *	–0.157	0.486
21	0.886	0.757	0.529	0.724
22	–0.057	0.929 *	0.157	0.343
23	0.500	0.457	–0.243	0.238
24	0.429	0.971 *	–0.300	0.367
25	0.700	0.929 *	–0.100	0.510
26	0.600	0.929 *	–0.443	0.362
27	0.500	0.814	–0.186	0.376
28	0.943 *	0.671	0.057	0.557
29	0.571	0.714	–0.071	0.405
30	0.629	0.771	–0.357	0.348
Aggregate r_s	0.754	0.886 *	–0.377	0.421

Note: Significance denoted by * is at the $p < 0.05$ level.

Table 2: Target-Judge agreement correlations