THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

# Rating e-mail personality at zero acquaintance

OPEN ACCESS

# 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*

**Abstract**

Electronic media play an ever-increasing role in our 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 produced by authors of known personality, who briefly described their recent activities, and were collected as part of a previously reported study which demonstrated linguistic characteristics of personality. As predicted by the perception literature, we find that even with minimal textual cues there is relatively high agreement, for ratings of author Extraversion. However, agreement for Neuroticism ratings appears to be further reduced by the environment, especially between target and judges. In addition to reducing the cues available for personality rating, the study extends the previous work in two main ways: first, it measures one further dimension of target personality—Psychoticism—rather than the separate factors Agreeableness and Conscientiousness (along with Openness); and secondly, it adopts additional, novel exemplar-based and subjective measures of personality perception. [Wordcount = 7,300]

## 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 via e-mail in particular? Although it is a written form of language, CMC is generally considered to share many similarities with spoken interaction (Bälter, 1998; Colley and Todd, 2002). Understanding how personality is projected and perceived through e-mail is a timely issue, given the continuing popularity of the medium (Baron, 1998).

In face-to-face interaction we are highly effective at judging people's personality (*e.g.*, Funder and Dobroth, 1987; Funder and Colvin, 1988; Paunonen, 1989), as well as other characteristics, such as familiarity, gender, emotion or temperament (*e.g.*, Cheng, O'Toole, and Abdi, 2001). However, e-mail is frequently used to make contact with people for the first time, while lacking many of the cues usually used for face-to-face personality judgement. Synchronic CMC environments are already known to have implications for personality judgement (Hancock and Dunham, 2001a; Markey and Wells, 2002), and so here we study the effects of asynchronous e-mail, upon person perception.

* Please address queries about this work to Alastair Gill or Jon Oberlander
  *Email addresses:* `A.Gill@ed.ac.uk` (Alastair Gill), `J.Oberlander@ed.ac.uk` (Jon Oberlander), `Elizabeth.Austin@ed.ac.uk` (Elizabeth Austin).

In pursuing this study, we take advantage of Eysenck's three-factor model of personality (Eysenck and Eysenck, 1991; Eysenck, Eysenck, and Barrett, 1985), with its dimensions of Extraversion, Neuroticism and Psychoticism; however, wherever relevant, we discuss connections to other work which has exploited the five-factor model (Digman, 1990; Costa and McCrae, 1992; Wiggins and Pincus, 1992; Goldberg, 1993), and its dimensions of Extraversion, Neuroticism, Conscientiousness, Agreeableness and Openness.

The rest of the paper is structured as follows. The next section provides general background to the study; we then outline the hypotheses tested, and detail the method and results, before concluding by way of discussion.

## 2 Background

This section provides general background on personality perception, discussing relevant work in terms of Funder's (1995) Realistic Accuracy Model. Then it briefly surveys results on the linguistic projection of personality, followed by work on the effects of computer-mediated communication both on language use, and on personality perception. The section concludes by framing the hypotheses to be investigated in the study.

### 2.1 Perception of Personality

Personality judgement data can be gathered in several ways. On the one hand, subjects' self-reports of personality, together with ratings of subjects by peers (such as spouses or colleagues), have been compared with each other for agreement. On the other hand, strangers have been called upon to make personal-

ity judgements, after being exposed to various different kinds of information about the target individuals. Research has investigated relevant factors which influence accuracy (Kenny, 1994; Funder, 1995), and here we adopt Funder's (1995) Realistic Accuracy Model to frame our study. This model views accuracy of judgement as a function of the relevancy, availability, detection, and utilisation of relevant behavioural cues. Furthermore, Funder outlines a 'path to accurate judgement', which grounds these processes in terms of the quality of the 'judge', 'target', 'trait', and 'information' in the study. We describe these in more detail below.

Good and bad judges are distinguished by their differing use of the cues which are available to them. For example, Funder (1995) proposes that knowledge about personality and the way it is revealed in behaviour would favour better socialised judges. This therefore implies that Extraverts make better judges than Introverts (low Extraverts), because they 'have more experience in social settings than introverts', and Funder cites studies which have shown this to hold for non-verbal cues in social interaction (Akert and Panter, 1988), and in determining the authenticity of suicide notes (Lester, 1991). Additionally, Funder also acknowledges the implications of judge ability and motivation; specifically, he notes the importance of the judge's intelligence, and the level of their commitment to the accuracy of their decision.

Good targets are proposed to be those whose behaviour makes available numerous and informative clues to their personality. In particular, Funder (1995) again notes the relevance of social behaviour—this time in the targets—since those with higher levels of social behaviours in particular exhibit more potential clues about their personality, relative to people who are less active (*e.g.* Borkenau and Liebler, 1992). Additionally, people who are high self-

4

monitors (Snyder, 1974, 1987) adjust their behaviour to changes in the social environment, and are predicted to be harder to judge accurately than low self-monitors, who are supposed to act consistently across different situations. Indeed, this raises difficulties parallel to those found in rating individuals with dishonest or socially undesirable behaviours. Those individuals are likely to try and conceal their behaviours, leading to difficulty in accurate judgement on the basis of their overt social behaviour (Funder, 1995).

Distinguishing between the different personality dimensions has shown that, even in judgements by close acquaintances, much greater agreement is found for ratings of Extraversion than for Neuroticism in both the EPQ (Gomà-i-Freixanet, 1997), and in the five factor models of personality (McCrae and Costa, 1987). For the EPQ, we find additionally that Psychoticism displays the lowest agreement in judgements and that additionally agreement for Lie-scale ratings is slightly higher than for Neuroticism (Gomà-i-Freixanet, 1997). For the other traits of the five-factor model, generally Openness shows similar levels of agreement to Extraversion, whereas Agreeableness shows low Agreement similar to that of Neuroticism, with Conscientiousness located somewhere between these groups (McCrae and Costa, 1987). Additionally, self-ratings were shown to be more informative in predicting behaviour for Extraversion than for Neuroticism (Spain, Eaton, and Funder, 2000).

Funder suggests that such findings are due to the different properties of traits, and has proposed that good traits are highly visible, and demonstrate low evaluativeness. Using Extraversion and Neuroticism as examples to which lay perceivers of personality show sensitivity, he 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

5

evaluativeness However, Neuroticism is lower in visibility (characterised by, *e.g.*, internal worrying thoughts or feelings), and is regarded as more 'evaluative', *i.e.*, affectively charged. It may thus 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 pass judgement on such behaviours, leading to reduced inter-judge agreement. When less evaluative measures of Neuroticism are used, agreement increases (John and Robbins, 1993).

The amount and relevance of target information available to the judges influences their agreement. Close acquaintances agree better with each other and with the target, than do relative strangers (Funder and Colvin, 1988; Paunonen, 1989; Paulhus and Bruce, 1992), although both predict target behaviour equally well, when they know the target in a relevant context (Colvin and Funder, 1991). Indeed, certain types of information can be more or less diagnostic of personality: for example, a person talking about their thoughts and feelings, rather than about hobbies, leads to more accurate judgement of their personality (Andersen, 1984), with similarly behaviour in unstructured situations being more informative than highly scripted tasks (Funder and Colvin, 1991).

Judgements by close acquaintances (especially when taken as a composite measure) generally also better predict target behaviour than judgements by other peers (Kolar, Funder, and Colvin, 1996). At the other extreme, studies have investigated personality perception of strangers on the basis of minimal cues, at so-called *zero-acquaintance*. Here there appears to be interaction between the available information and the visibility of the trait being judged. This has been demonstrated using solely linguistic or visual cues. From ex-

posure to transcribed interactions, self-other, or *target-rater*, agreement was shown for ratings of Extraversion and Introversion (Gifford and Hine, 1994). Alternatively, Albright, Kenny, and Malloy (1988) found that, on the basis of physical appearance, Extraversion and Conscientiousness could be reliably rated, however, the judgements of Extraversion appeared to be mediated—or influenced—by judgements of the physical attractiveness of the target. Judgements made at zero-acquaintance appear readily influenced by stereotypes, which judges may attend to in the absence of readily available cues. For example, perceptions of target nationality or gender (Gallois and Callan, 1986) may influence accuracy, in addition to ideas about personality (McCrae and Costa, 1987).

## 2.2   Personality and Language

With this background on perception of personality in place, we now turn to work on the relations between personality and language; the greater part of this work has been on projection, rather than perception, and what follows is covered in greater detail in Oberlander and Gill (2004). It is notable that the majority of work to date has focussed on speech rather than writing, and the emphasis has been on Extraversion, and to a lesser extent Neuroticism. Given that our focus is on written e-mail, we do not discuss features specific to speech (such as speech rate), but focus instead on grammar and lexical content, whether in spoken or written language. Reviews of various aspects of this work can be found in Scherer (1979), Furnham (1990), Smith (1992), Dewaele and Furnham (1999), and Pennebaker and King (1999).

Furnham (1990) proposed that Extravert speech and language use would have a number of distinctive features. Speech would tend towards non-standard accents, a higher speech rate, and more dysfluencies. Language would be less formal, with a more restricted code, with loose use of vocabulary, and it would be more implicit. This is defined as a showing a preference for pronouns, adverbs and verbs, and a dispreference for nouns, modifiers and prepositions. Heylighen and Dewaele (2002) argue that Extraverts produce implicit (or informal) language because it requires less cognitive effort, and relies more on the context for interpretation. Extravert language—in both native and non-native speakers—has been shown to be more implicit, and to possess higher numbers of words (Furnham, 1990; Dewaele and Furnham, 1999, 2000; Dewaele, 2001). Heylighen and Dewaele (2002) note that Introvert language features tend to be closely related to those of formal language, and this is consistent with the finding that Extraverts demonstrate lower lexical richness in formal situations (Cope, 1969; Dewaele, 1993; Dewaele and Furnham, 2000).

Turning to lexical biases, we focus on results which have been obtained using the Linguistic Inquiry and Word Count text analysis program (LIWC; Pennebaker and Francis, 1999, see also the more recent LIWC2001; Pennebaker, Francis, and Booth, 2001). Pennebaker and King (1999) applied LIWC analysis to texts written by authors for whom (five-factor) personality information was available, and uncovered a number of significant patterns. For instance, High Extraverts use more social process and positive emotion words. High Neurotics use more first person singular and negative emotion words. High Openness scorers use more articles, longer words and insight words. High Agreeableness scorers use more first person singular and positive emotion words. High Conscientiousness scorers use more positive emotion words. In

8

a perception study based on LIWC, Berry, Pennebaker, Mueller, and Hiller (1997) found that transcribed texts rated as higher in Dominance used fewer positive emotion words and self referents, and texts regarded as displaying greater Competence used fewer self referents and negations, and more present tense verbs. Texts regarded as displaying greater Dominance and Competence were longer (these texts also show lower lexical diversity, but this appears to be a length effect, see Gill, 1998, for a discussion).

## 2.3 Computer-mediated communication

As we noted in the introduction, computer-mediated communication is likely to have effects on the ways in which people express themselves, and hence on how others perceive them. This subsection therefore discusses each of these points in turn, and then summarises the hypotheses to be tested in the study.

### 2.3.1 Effects of CMC on language

Computer-mediated communication, and more specifically e-mail, is considered to be a form of communication located between the domains of speech and writing: it shares properties of both modalities (Bälter, 1998; Baron, 2001). For example, it is a written form with interlocutors physically separated, and 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. Additionally it has also developed its own stylistic features (Baron, 1998). Colley and Todd (2002) refer

to stylistic "emailisms" described by Petrie[1] which are common to e-mail, but rare in other forms of writing. These include trailing dots, capitalisation, excessive use of exclamation marks and question marks; but use of 'emoticons' was found to be rare. Study of a bulletin board corpus (e-mails posted to the web) using a multi-dimensional analysis similar to that of Biber (1993), found that the language genre was most like that of 'public interviews and letters, personal as well as professional' (Collot and Belmore, 1996).

Computer-mediated communication provides impoverished cues, and is less rich than face-to-face communication (Panteli, 2002); thus, information has to be communicated using alternative means. Werry (1996) notes that in internet relay chat (interactive electronic communication) innovative linguistic strategies are adopted to represent the intonational or paralinguistic features of face-to-face discourse, with this finding mirrored in coordination devices employed in task-based interaction in a CMC environment (Hancock and Dunham, 2001b).

Although CMC lacks cues compared to face-to-face interaction, it still provides rich information about the communicator. For example Panteli (2002) found that the construction of text-based messages conveyed the social cues indicating status differences in organisations. Additionally, several studies have shown gender to be communicated in a CMC environment: in mailing lists, messages written by females used more interactional features, and communicated more information, whereas males were more critical (Herring, 1996); in e-mails to friends, females preferred social and domestic topics, whereas males

---

[1] The study which Colley and Todd (2002) refer to was published on-line, and downloaded by them in 2000; however, the link they publish is no longer available.

preferred impersonal and external topics (Colley and Todd, 2002); interlocutors and judges were consistently able to identify author gender from e-mails, with female messages found to be characterised by more modal auxiliaries, intensifying adverbs, mention of emotions, sharing of personal information, questions, compliments, apologies, and self-derogatory remarks. Conversely, males were found to give more opinions and use more insults (Thomson and Murachver, 2001). In addition, style matching was found for interlocutors of the minority gender style when communicating with those belonging to the norm group, regardless of their own gender (Herring, 1996; Thomson, Murachver, and Green, 2001).

Corpus-based work on e-mail, which links language with personality projection, has found that Extraversion, Neuroticism and Psychoticism can each explain between 11% and 27% of language variation found using standard text analysis techniques (Gill, 2003). Furthermore, this work has both confirmed a number of general findings, and uncovered relations between specific emailisms and personality features (Gill and Oberlander, 2002; Oberlander and Gill, 2004). On the one hand, it was found that: Extraversion is associated with fluency, positivity and implicitness; Neuroticism with self-concern, negativity and implicitness; and Psychoticism with creativity and detachment. On the other hand, a number of emailisms were shown to be associated with particular personality subgroups. For instance, high Extraverts were heavy users of ellipses (trailing dots), and high Neurotics were heavy users of multiple exclamation marks.

Additional properties of the CMC environment are that it enables and encourages increased communication. For example, in computer-mediated task-based group meetings, low Extraverts provided more original solutions than in the

face-to-face meetings (although in the latter environment they provided more comments). In each case high Extraverts showed greater participation in both environments than the low Extraverts (Yellen, Winniford, and Sanford, 1995). This fits with the earlier finding that Extraverts show greater desire to communicate and initiate interactions in face-to-face situations (McCroskey and Richmond, 1990). The pattern of CMC behaviour is mirrored with second language learners, with students who are less forthcoming in class being more inclined to contact their teacher by e-mail (Bloch, 2002).

### 2.3.2 Effects of CMC on personality judgement

When the availability of infomation for personality judgements is reduced, we find that accuracy is also reduced. For example, judges who are better acquainted with the target generally provide more accurate personality ratings, as discussed above (section 2.1). Whether or not subject and judge have prior knowledge of each other, technology also has an impact on what information is available in a communicative situation. Zero-acquaintance judgements are particularly vulnerable to technological artifacts. For example, interviews conducted over the telephone were found to result in reduced self-interviewer and peer-interviewer agreement than face-to-face interviews (Blackman, 2002). Furthermore in text-based computer-mediated environment (CMC) judgements of gender, accuracy was reduced by expectations of linguistic stereotypes for the male and female writers (Savicki, Kelley, and Oesterreich, 1999). For judgements of personality in CMC (following one-on-one interactions in an internet chat room), consensus was found between judges for a target's Extraversion, Agreeableness, and Openness, but target-judge agreement was only found for Extraversion and Openness (Markey and Wells, 2002).

12

Impressions of personality formed following task-oriented synchronous computer-mediated communication found that they were less detailed but more intense compared with those from face-to-face communication. Specifically, in the CMC environment, judges seemed less able to rate their partners for Extraversion, Neuroticism, and Agreeableness. Across both environments, Conscientiousness, Agreeableness, and Extraversion were the most rateable (Hancock and Dunham, 2001a).

## 2.4 Perception Hypotheses

On the basis of previous perception studies and the properties of the traits themselves, we set out to test the following hypotheses:

**Psychoticism** We expect that agreement will be lowest for Psychoticism, due to its high evaluativeness and lower visibility, which we predict will be most affected by the lack of information available in the CMC and zero-acquaintance conditions.

**Extraversion** This trait will be the most easily perceived due to its high visibility and low evaluativeness. We therefore expect it to show the highest levels of inter-judge and target-judge agreement, even in CMC at zero-acquaintance.

**Neuroticism** We propose that Neuroticism is less visible than Extraversion and less evaluative than Psychoticism. So we expect agreement to be higher than for Psychoticism, but lower than for Extraversion. We also expect that the CMC and zero-acquaintance conditions will only have a moderate lowering effect upon agreement.

# 3    Method

## 3.1    The 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 (rating themselves between 7 and 10 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 personality, although 9 had 'some' experience—having read books on psychology—and 3 had studied psychology or personality psychology as part of their degree). No one 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.

### 3.2.1    The target texts

The target texts were all taken from e-mail data collected previously. The corpus consisted of 210 texts (2 texts each from 105 subjects, who had also provided EPQ-R short form data). The texts had been written under experimental conditions, as if to a good friend; for each subject, one text informed

14

the friend of the subject's activities over the previous week, and the other related plans for the next week. From this set of 210 texts, 18 were selected for the experiment: for each of the three-factor personality dimensions of Psychoticism, Extraversion and Neuroticism, 6 texts were chosen to represent a range of scores. For each dimension, two texts were chosen whose authors scored greater than $+1$ standard deviation from the mean for that personality dimension, and two were chosen which were greater than $-1$ standard deviation. In each case, these texts scored less than 1 SD either side of the mean on the other personality dimensions. Two further texts were selected which were within 1 SD, but more than .5 SD of the mean: one each above and below the mean. In these cases, the texts were within 1 SD (.5 SD where possible) of the mean on the other personality dimensions.

In this experiment, texts detailing 'past' activities were selected for the rating exercise as these were generally longer than those outlining future plans (Mean length in words: Psychoticism texts=258.67, Extraversion texts=261.33, Neuroticism texts=261.00).These selected texts were presented in random order of personality score for each dimension at a time.

### 3.2.2   The questionnaire

The rating questionnaire was divided into three sections, each relating to a different personality trait (Psychoticism, Extraversion, or Neuroticism). [2] The order in which these sections were presented was determined by a Latin square technique to avoid an ordering effect. These booklets were given an identifica-

---

[2]  Note, however, that the terms Tough-mindedness and Emotionality were used instead of Psychoticism and Neuroticism; see also below for further details.

tion code which was used when referring to judges in order to maintain their anonymity.

The rating questionnaire booklet was prefixed by an explanatory page informing judges of the format of the experiment, and emphasising our interest in how they 'think the author comes across', the need for them to answer 'honestly and accurately' and 'not to spend too long thinking about each question' and to instead concentrate on giving their 'initial response'. For each personality dimension a description based upon those of Eysenck and Eysenck (1975) was included. These descriptions received minor re-wording to enhance intelligibility, minimise issues of social desirability, and to make them more understandable to a wider audience (as recommended by Eysenck and Eysenck, 1975, p. 12). Although it is more usual to rate personality using a standard set of questions (Eysenck *et al.*, 1985; Costa and McCrae, 1992; cf. Ten-Item Personality Inventory, Gosling, Rentfrow, and Swann, 2003), Sneed, McCrae, and Funder (1998) have found that 'most laypersons can easily grasp the nature of the factors and their behavioural manifestations and can spontaneously recognise their grouping when presented with clear exemplars' (p. 115).

Judges were at first asked to rate the personality of the author for the trait which has been described at the beginning of the section, using the following question 'How [Tough-Minded/Extravert/Emotionally-Stable] is the the author of the e-mail', with the extremes of the scale labelled 'Not at All' and 'Very [Tough-Minded/Extravert/Emotionally-Stable]'. The judges were then asked 'How easy was it to come to this conclusion?' (about the e-mail author's personality) rated on a scale of 1–10 labelled 'Very Difficult' and 'Very Easy' respectively.

*3.3 Procedure*

All 30 judges worked through the rating booklet at their own speed, and although there was no official time limit, they were encouraged to work 'quickly and efficiently' so that the participant did not spend too much time thinking about their responses and also so that they remained well motivated. In all cases several judges participated in the experiment at the same time, over-seen by the experimenter. However, they were informed that exam-type conditions should be maintained, and that responses to the questionnaire should not be discussed with each other during the experiment. Equal numbers of participants were randomly assigned to each questionnaire (these questions are detailed above). After completing the rating booklet, there was a short debriefing session which included administration of the EPQ-R (Eysenck *et al.*, 1985).

# 4   Results

*4.1   Judges*

The judges' completion of the of EPQ-R (short form, Eysenck *et al.*, 1985) 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. The judges' personality profile is therefore similar to the published norms.

| 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

*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. The value of these coefficients were: Psychoticism 0.287 $[W(5) = 43.05, p < 0.0001]$; Extraversion 0.471 $[W(5) = 70.64, p < 0.0001]$; Neuroticism 0.266 $[W(5) = 38.91, p < 0.0001]$.

In addition to using Kendall's *W* coefficient of concordance which describes judge consistency overall, it is also possible to examine how the each judge agrees with each of the other judges in the experiment (cf. Morris, Gale, and Duffy, 2002). Correlations were performed for each judge with each of the other judges, with the mean overall correlation reported for each judge (counts of correlations achieving significance are also noted for each cell out of a maximum of 29). Although the personality questionnaire results can usually be regarded as interval data (Kline, 1983), the ordinal nature of the rating scale responses meant that Spearman rank correlations were used throughout the following analyses, since this is more appropriate for such data (Butler, 1985).

The final row of Table 1 gives the average rank correlations for each trait across all judges. Extraversion is shown to have the greatest inter-judge agreement, and therefore in terms of inter-judge agreement appears to be the easiest trait to rate (mean $r_s = 0.482$). This is followed by Psychoticism (mean $r_s = 0.333$), and finally Neuroticism (mean $r_s = 0.308$) which both show lower levels of agreement. This therefore suggests that they are harder to rate. The greater

agreement shown between judges for ratings of Extraversion is also reflected in the total number of significant correlations found for the trait (76), which is much greater than that found for either Psychoticism (26) or Neuroticism (20).

Since we calculate Spearman rank correlations, here we have reported the means of these correlations (Morris *et al.*, 2002), rather than use Fischer's *r* to *z* conversion (*e.g.*, Funder and Colvin, 1988; Funder, Kolar, and Blackman, 1995; Vogt and Colvin, 2003). In order in order 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, since this statistic is regarded as the equivalent of performing correlations between all possible pairs of raters (McCrae and Costa, 1987). Similarly to the findings reported in Table 1, Extraversion showed the highest agreement with an intraclass correlation of 0.403, and although Neuroticism and Psychoticism both showed relatively low agreement, this was actually slightly lower for Psychoticism (0.206) than for Neuroticism (0.248; all significant at p< 0.0001).

*4.3   Are All Judges Equally Good?*

The level of agreement between judges across all three personality traits is also shown in Table 1. From this it can be seen that the best judges, *in terms of agreeing most with the others* were judges 20, 30, 7, 28, and 29, and the worst judges were 5, 19, and 1. The mean level of agreement across P, E, and N dimensions was 0.374.

Turning to each trait individually, for Psychoticism judges 15, 4, 6, 30, and 16

showed the most agreement, whilst judges 5, 13, 19, 24, and 3 showed relatively little agreement. For Extraversion, judges 7, 24, 11, 26, and 20 demonstrated greatest agreement, whereas for judges 5, 14, and 1 the levels achieved were much lower. For Neuroticism it can be seen that judges 14, 20, 13, and 5 all show the most agreement, whereas judges 27, 19, and 1 actually show disagreement with other judges.

The level of agreement between target and judge ratings can also indicate how accurate judges are, and information about this can be found in Table 2. Here it can be seen that the best judges *in terms of agreeing most with targets across all personality dimensions* are judges 21, 17, 6, 11, 18, and 28 and the worst judges are 8 and 12 who both correlated negatively, and judges 13 and 5.

For each individual trait, starting with Psychoticism, judges 28, 14, 21, and 17 all agreed highly with the targets, whereas judges 5, 3, and 22 showed a negative correlation with the target self reports of personality. The trait of Extraversion elicited even higher levels of target-judge agreement for judges 20, 22, 25, and 26, with only judges 1 and 8 showing a negative correlation. However, for Neuroticism lower levels of agreement were found for judges 18 and 21, with many judges showing a negative correlation (16 in total), with some of the greatest disagreement found for judges 12 and 13. Additionally we analysed inter-judge and judge-target agreement by the personality traits of the judges (EPQ-R and NEO-PI-R), but this appeared to demonstrate little effect on levels of agreement.

## 4.4  Are All Targets Equally Good?

If one text on a particular personality trait was much more difficult to rate than any of the others, we would expect judges to show a much greater variability in their ratings for it. Levene's test for homogeneity (or equality) of variance was used to investigate whether there was significant variance in ratings for texts belonging to each trait. Although significant differences were not found for Extraversion or Neuroticism, they were found for Psychoticism [$F(5, 174) = 2.868, p < 0.05$]. In this case, the texts which showed the greatest variance were P6 (M=4.4, SD=2.3; mid-high-P), P5 (M=4.4, SD=2.0; high-P), and P3 (M=5.2, SD=2.0; high-P), and therefore appear to be the most difficult to rate. The texts showing least variance were P4 (M=2.7, SD=1.4; mid-low-P), P1 (M=2.8, SD=1.6; low-P), and P2 (M=3.5, SD=1.9; low-P). This demonstrates that the high Psychotic texts showed greater variation in ratings, and may indicate that they were harder to rate, therefore resulting in the lower intraclass correlation results for ratings of Psychoticism.

## 4.5  Target-Judge Correlation

To gain an overall sense of how the individual judges had performed, mean correlations of judge-target agreement were calculated. For each of the judges, each of their six ratings of the texts for P, E, and N were correlated with the original personality scores of the authors, and their mean performance for rating P, E, and N also noted (Table 2). Looking at the correlations of the individual judges for each dimension, we can see that the largest number of significant correlations (out of a possible 30) were found for Extraversion (5),

| 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

followed by Psychoticism (2), with none of the correlations between judges and targets reaching significance for ratings of Neuroticism.

To ensure increased agreement and accuracy of target-judge correlation, the aggregate measure of personality ratings across multiple raters was then calculated, since McCrae and Costa (1987) suggest that this takes into account how the target is seen by the judgement group as a whole. Therefore Spearman correlations were performed taking the mean of the judges ratings for each text, along with the original personality scores of the targets. Correlation of the target's raw EPQ-R with the mean of the judges ratings (1–10), gave the following correlations (Spearman, pairwise, two-tailed, 6 cases): Extraversion $r_s = 0.886$; Psychoticism $r_s = 0.753$; Neuroticism $r_s = -0.377$. Of these, only ratings of Extraversion showed significant target-judge agreement ($p < 0.05$; Psychoticism demonstrated a lower level of significance at $p < 0.1$).

## 4.6   Judge Perception of Target Rating

### 4.6.1   Perceived similarity of target-judge

In order to investigate how judges perceived the target author personalities relative to their own, analysis of the similarity ratings of texts was performed.

These analyses were carried out with the six target texts for each personality dimension grouped into three 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 also this time for Extraversion [$F(2, 58) = 4.052, p < 0.05, MSE = 1.6$], but not Neuroticism texts. Tukey HSD tests revealed that significant dif-

ferences 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 $p < 0.05$).

These analyses have so far not taken into account the effects of judge personality on the ratings of similarity, but have grouped the judges as a whole. Therefore, judges were categorised as either 'high' or 'low' on the personality dimension in question using a mean split, and author personality of the target texts was categorised into the High, Mid, or Low groups, since this reduced the data yet retained broad information. 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. For Extraversion, as expected, a main effect was found 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 [$F(2, 56) = 3.430, p < 0.05, MSE = 1.5$]. No effects were found for Neuroticism.

In order to investigate possible interaction effects further, we examine the simple main effects of text author personality for the high and low personality groups of judges individually. The within subjects ANOVA shows—as expected from the significant interaction—effects of text type on the ratings of similarity for high Extravert judges [$F(2, 26) = 5.082, p < 0.05, MSE = 1.9$]. 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 ($M = 4.5$).

However, findings for Psychoticism also show an effect of text type on similarity rating for the low Psychotic judges [$F(2, 32) = 5.753, p < 0.01, MSE = 1.9$]. 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 ($M = 4.6$). For high Psychotic judges, MidP texts were regarded as most similar ($M = 4.8$), and HighP texts most dissimilar ($M = 3.8$), but this effect of text type was found to be border line significant at $p < 0.1$ [$F(2, 24) = 3.299, p < 0.1, MSE = 1.2$]. No significant effects were found for judges grouped by Neuroticism.

If the actual personality scores of the texts being rated for similarity are disregarded, and the personality scores of the raters are considered (again divided at the mean as either high or low), then between subjects ANOVA shows that only rater Psychoticism has an influence on ratings of Psychotic texts [$F(1, 28) = 6.556, p < 0.05, MSE = 1.0$]. This means that low Psychotic judges rated the texts (*all texts*, High *and* Low P) as more similar ($M = 5.4$) than high Psychotic judges ($M = 4.4$) ($p < .05$).

### 4.6.2 Perceived ease of rating personality

Indications of how judges perceived ease of rating personality of texts were gained from the subjective scores. Within subjects ANOVAs were performed for ratings of ease compared with the personality of the text author categorised into High, Mid and Low. ANOVAs show that significant effects of the personality of the text upon rating difficulty for Extraversion [$F(2, 58) = 13.155, p < 0.001, MSE = 3.2$] and Psychoticism [$F(2, 58) = 10.368, p < 0.001, MSE =$

1.9]. Tukey tests show that significant differences for Extraversion exist between LowE ($M = 5.7$) and HighE ($M = 7.8$), and between HighE ($M = 7.8$) and MidE ($M = 5.8$) texts, and for Psychoticism between HighP ($M = 5.9$) and LowP ($M = 7.4$), and between HighP ($M = 5.9$) and MidP ($M = 7.1$) texts (all $p < 0.05$). For completeness, the above analysis was repeated taking into consideration the personality of the judges. These analyses confirmed the above findings, and as expected, no main effects of judge personality or interaction effects of judge personality and text author personality were found for any of the traits.

## 5  Discussion

According to our hypotheses (stated in Section 2.4), 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 perception would be worst for Psychoticism, and best for Extraversion, with Neuroticism in the middle. Broadly, it appears that the Extraversion hypothesis holds, but the others do not. To see that this is so, we discuss the range of results in more detail.

### 5.1  Ratings of Inter-Judge and Target-Judge Agreement

The results demonstrate that judges reliably agree with each other when rating a text for a specific personality trait. However the level of agreement is greatest for Extraversion, followed to a lesser extent by Psychoticism and then Neuroticism. That Extraversion showed highest agreement was predicted in

27

our hypotheses; however, that Neuroticism showed the lowest agreement was not. How did this come about?

Extraversion shows the greatest inter-judge agreement. This is consistent with previous findings, and may be due to Extraversion's more observable and less evaluative properties (noted in Section 2.1). However, in the case of John and Robbins's analysis, Neuroticism (termed Emotional Stability in their model; along with Intellect, or Openness to Experience) shows quite good agreement, with this reduced for Conscientiousness and lower still for Agreeableness (John and Robbins, 1993). In the present study, because the three factor (EPQ-R) personality model was used, Psychoticism has replaced Conscientiousness and Agreeableness traits (whilst Intellect is not measured). This has left Neuroticism as the trait showing least inter-judge reliability. We are comparing two different models of personality; hence, it is difficult to assess whether in the current study Neuroticism has been shown to demonstrate less agreement in judges than in previous studies, or whether in fact Psychoticism is more observable and less evaluative than the individual traits of Conscientiousness and Agreeableness.

However, since the actual ratings in the current study are using a different novel source of information as the target (a short sample of e-mail text rather than having met the person in real life or through observation; cf. Markey and Wells, 2002, who used an interactive CMC chatroom environment), this difference in rating agreement, for both Neuroticism (Emotional Stability) and Psychoticism (Intellect/Conscientiousness/Agreeableness) may be due to the properties of e-mail text as not being 'good information' for personality judgement of Neuroticism (Funder, 1995).

Turning to the agreement between the judges' and targets' rating of personality, and a similar pattern emerges to that of inter-judge agreement, with ratings for both Extraversion and Psychoticism showing a relatively stronger positive correlation, but Neuroticism bearing a non-significant negative relationship. This again points to Extraversion being an observable, but relatively unevaluative trait, its evaluative neutrality emphasised by self-peer agreement. The weaker target-judge relationship for Psychoticism would suggest that it is both less observable and more evaluative. However the lack of strong target-judge relationship for Neuroticism relative to Psychoticism (cf. Gomà-i-Freixanet (1997), and given their similar inter-judge agreement) would suggest the high Neurotic individual's awareness of the trait's evaluativeness results in a distortion of self-perception, or alternatively that e-mail does not provide good information for its accurate judgement.

However, as previous studies of the e-mail data have shown, there are linguistic features of Neuroticism (Gill, 2003; Oberlander and Gill, 2004). Therefore, one possibility is that the judges are attending to the *wrong* information. In a study which looked at personality perception through speech, Scherer (1972) found that despite the high rate of inter-rater reliability for the trait of Extraversion, there was little target-judge agreement for this trait. He concluded from this that judges were instead attending to stereotyped cue information for socially desirable traits projected by the targets. By analogy, in our case of Neuroticism, judges may be attending to misinformed stereotyped cues. However, given that Neuroticism is generally regarded as more evaluative and less desirable, it is possible that they attend to less desirable stereotyped features.

When the performance of individual judges is examined, it can be seen that inter-judge agreement can be differentiated across the traits: on some, this can

29

be quite high, and on others—especially Neuroticism—this can be quite low. In the case of target-judge agreement, the pattern is more consistent, with judges showing either generally higher or lower levels of agreement across all traits. This greater consistency of agreement is to be expected—due to a judge's ratings only being correlated with those of the target, rather than all of the other judges. As expected from the mean ratings of judges overall, most judges show a noticeably poorer performance for Neuroticism.

## 5.2  *Judge Perception Rating Measures*

We also collected novel subjective ratings of similarity between rater and target, and perceived ease of rating the text for personality. This data is informative because it allows us investigate how perceptions of the rating exercise and of own and other personality compare to objective measures.

For the similarity ratings there was a general pattern of the judges distancing themselves from the undesirable end of the trait. Even when judge personality was taken into consideration, the judges were still seen to identify with low Psychoticism, meaning that, whilst the low Psychotic judges (accurately) rated the low Psychotic texts as most similar, the high Psychotic judges also (incorrectly) rated the low Psychotic texts as most similar.

Although it may be the case that highly Psychotic judges are for some reason less able to accurately judge author Psychoticism, it would appear to be more likely that they were influenced by the evaluativeness of this trait. Indeed, it may be that as a result of higher levels of judge Psychoticism, such judges are more likely to consciously or unconsciously provide inaccurate information

about themselves. Level of judge Psychoticism also had an overall effect on the similarity scores, with low Psychotic judges regarding themselves as more similar in general to the authors of the texts. Given that lower Psychoticism scorers are more likely to be interpersonally oriented it should not be too surprising that they more readily identify with the authors of the texts, regardless of how similar their personality scores actually were.

For the judges of Extraversion overall, a relationship was only shown between the texts when grouped into three categories, with the high Extravert text regarded as more similar than the mid text. When personality information is added to this analysis, an interaction effect emerges between the personality of the judge and the author of the text. Separate analysis of the high and low Extravert similarity ratings shows that the high Extraverts view the high Extravert texts as most similar by quite some way (followed, surprisingly, by the low Extravert texts). On the other hand, this interaction is mirrored by low Extravert judges (not significantly) rating the low Extravert texts as most similar, followed shortly after by the high Extravert texts. Since both groups accurately rate the texts which are most similar to themselves, this contributes to the interaction effect. However because low Extravert judges rate the high Extravert texts as still relatively similar, this contributes to the overall effect for high Extravert texts being rated as similar for the group as a whole.

Since an interaction of judge and author personality occurs, this suggests that effects of trait desirability, or undesirability, are less important for ratings of Extraversion, and this is consistent with its being considered a less evaluative trait. Furthermore, the fact that high Extraverts identified their similarity more accurately may be a result of their greater interpersonal ability associated with higher Extraversion. However, the fact that low Extraverts are less likely

to distinguish themselves as low Extraverts as opposed to high Extraverts may be an effect of a lower interpersonal awareness or a remnant of weak desirability effects of higher levels of Extraversion.

So far we have discussed the accuracy and relative desirability effects present in the similarity ratings for Psychoticism and Extraversion, without reference to Neuroticism. Whilst Psychoticism and Extraversion have shown several broad patterns relating to similarity ratings, perception of similarity to Neuroticism show few patterns and again demonstrates a mixed picture.

Turning to the perceived ease of rating texts, it is apparent that these findings are consistent across both Extraversion and Psychoticism dimensions: high Extravert texts and low Psychotic texts are regarded as the easiest to rate, regardless of judge personality. These findings are consistent in that they show that texts belonging to the more desirable end of both scales, are seen as easier to rate. The results are therefore not an artifact of the rating scale description (in which case the higher ends of the scales would have been regarded as easier to rate). So we suggest that it may be that individuals have a clearer concept of behaviour which is desirable, rather than undesirable.

## 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 (somewhat surprisingly) to a slightly lesser extent, for Psychoticism. In both cases, judges used an exemplar-based rating of personality rather than an itemised

personality questionnaire. Additionally, judges rated ease of assigning personality and also perceived target similarity, which confirmed the judges' ability to perceive personality consciously and subconsciously, which provided further information about the relative evaluativeness and desirability of these traits.

Although judges generally agreed with each other regarding ratings of Neuroticism, unexpectedly little consistency was found with the author's own personality assessment, or with ease of rating or similarity. We propose that this is partly due to characteristics of the trait itself, and also the quantity and quality of information which the e-mails in this experiment made available to the judges.

## 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.

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

Bälter, O. (1998). *Electronic Mail in a Working Context*. Ph.D. thesis, Royal Institute of Technology, Stockholm.

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.

Berry, D., Pennebaker, J., Mueller, J., and Hiller, W. (1997). Linguistic bases of social perception. *Personality and Social Psychology Bulletin*, **23**, 526–537.

Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, **19**, 219–241.

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.

Butler, C. (1985). *Statistics in Linguistics*. Blackwell, Oxford.

Cheng, Y., O'Toole, A., and Abdi, H. (2001). Classifying adults' and children's faces by sex: Computational investigations of subcategorial feature encoding. *Cognitive Science*, **25**, 819–838.

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, editor, *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, pages 13–28. Benjamins, Amsterdam.

Colvin, C. and Funder, D. (1991). Predicting personality and behaviour: A boundary on the acquaintance effect. *Journal of Personality and Social*

*Psychology*, **60**, 884–894.

Cope, C. (1969). Linguistic structure and personality development. *Journal of Counselling Psychology*, **16**, 1–19.

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

Dewaele, J.-M. (1993). Extraversion et richesse lexicale dans deux styles d'interlangue française [Extraversion and lexical richness in 2 styles of French interlanguage]. *I.T.L Review of Applied Linguistics*, **22**, 87–105.

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

Dewaele, J.-M. and Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning*, **49**, 509–544.

Dewaele, J.-M. and Furnham, A. (2000). Personality and speech production: a pilot study of second language learners. *Personality and Individual Differences*, **28**, 355–365.

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. (1975). *The Eysenck Personality Questionnaire*. Hodder and Stoughton, London.

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

Eysenck, S., Eysenck, H., and Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, **6**, 21–29.

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 Colvin, R. (1991). Explorations in behavioural consistency: Properties of persons, situations, and behaviours. *Journal of Personality and Social Psychology*, **60**, 773–794.

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., Kolar, D., and Blackman, M. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, **69**, 656–672.

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, editors, *Handbook of Language and Social Psychology*, pages 73–95. Wiley, Chichester.

Gallois, C. and Callan, V. (1986). Decoding emotional messages: Influences of ethnicity, sex, message type, and channel. *Handbook of Language 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*. Ph.D. 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*

36

*Science Society*, pages 363–368.

Gill, A. J. (1998). *Type-Token Ratio and the Measurement of Competence in Task-Oriented Dialogue*. Master's thesis, School of Cognitive Science, University of Edinburgh.

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

Gomà-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. (2001a). Impression formation in computer-mediated communication. *Communication Research*, **28**, 325–347.

Hancock, J. and Dunham, P. (2001b). Language use in computer-mediated communication: The role of coordination devices. *Discourse Processes*, **31**, 91–110.

Herring, S. (1996). Two variants of an electronic message schema. In S. Herring, editor, *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, pages 81–106. Benjamins, Amsterdam.

Heylighen, F. and Dewaele, J.-M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science*, **7**, 293–340.

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*. Guilford, New York.

Kline, P. (1983). *Personality: Measurement and Theory*. Hutchinson, London.

Kolar, D., Funder, D., and Colvin, C. (1996). Comparing the accuracy of personality judgements by the self and knowledgeable others. *Journal of Personality*, **64**, 311–337.

Lester, D. (1991). Accuracy of recognition of genuine versus simulated suicide notes. *Personality and Individual Differences*, **12**, 765–766.

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.

McCroskey, J. and Richmond, V. (1990). Willingness to communicate: A cognitive view. *Journal of Social Behaviour and Personality*, **5**, 19–37.

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

Oberlander, J. and Gill, A. (2004). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Submitted*.

Panteli, N. (2002). Richness, power cues and email text. *Information & 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 Francis, M. (1999). *Linguistic Inquiry and Word Count*

(LIWC). Lawrence Erlbaum Associates, Mahwah, NJ.

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

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC2001)*. Lawrence Erlbaum Associates, Mahwah, NJ.

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, editors, *Social Markers in Speech*, pages 147–209. Cambridge University Press, Cambridge.

Smith, C. (1992). Introduction: inferences from verbal material. In C. Smith, editor, *Motivation and personality: Handbook of thematic content analysis*, pages 1–17. Cambridge University Press, Cambridge.

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.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, **30**, 526–537.

Snyder, M. (1987). *Public appearances, private realities: The psychology of self monitoring*. Freeman, New York.

Spain, J., Eaton, L., and Funder, D. (2000). Perspectives on personality: The relative accuracy of self versus others for the prediction of emotion and

behaviour. *Journal of Personality*, **68**, 837–867.

Thomson, R. and Murachver, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology*, **40**, 193–208.

Thomson, R., Murachver, T., and Green, J. (2001). Where is the gender in gendered discourse? *Psychological Science*, **12**, 171–175.

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

Werry, C. (1996). Linguistic and interactional features of internet relay chat. In S. Herring, editor, *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, pages 47–63. Benjamins, Amsterdam.

Wiggins, J. and Pincus, A. (1992). Personality: Structure and assessment. *Annual Review of Psychology*, **43**, 473–504.

Yellen, R., Winniford, M., and Sanford, C. (1995). Extraversion and introversion in electronically-supported meetings. *Information & Management*, **28**, 63–74.