Running Head: CONTEXT, JUDGEMENT PERSPECTIVE, CUES AND ACCURACY

Looking at the same interaction and seeing something different: The role of informational contexts, judgement perspective and behavioural coding on judgement accuracy

Helen J. Wall*[1] Paul J. Taylor[2] Claire Campbell[3] Derek Heim[1], & Beth Richardson[4]

[1] Edge Hill University, Lancashire, UK; [2] Lancaster University, Lancashire, UK; [3]Ulster University, [4] University of Central Lancashire

***Corresponding Author at**:
Department of Psychology, Edge Hill University, St Helen's Road, Ormskirk, Lancashire, UK, L39 4QP
Tel: +44 1695 657336 Email: Helen.Wall@edgehill.ac.uk

Looking at the same interaction and seeing something different: The role of information, judgment perspective and behavioural coding on judgment 'accuracy'

We often make judgments about others' personality based on limited informational cues and in varying contexts, including interacting with a person or observing them from afar. The validity of such first impressions can have far-reaching outcomes for the success of future interactions, the types of people we choose to befriend and trust, and even the effectiveness of therapeutic interventions (see Funder, 1999). Research indicates that the validity or 'accuracy[1]' (i.e., similarity between an individual's personality score and another's perception of that individuals personality) of initial judgments can be high, which is surprising given the limited cues that are available in first impression encounters. Although recent studies have demonstrated the complexity in judgment accuracy in terms of moderators such as interaction demands (Wall, Taylor, & Campbell, 2016) and context type (Gosling, Ko, Mannarelli, & Morris, 2002; Letzring, Funder, & Wells, 2006; Wall, Taylor, Dixon, Conchie, & Ellis, 2013; see Funder 1999 for detailed overview) further exploration into when and how people form judgements of another's personality is warranted. In support, Back and Nestler (2016) have recently suggested that accuracy research can benefit from a focus on more complex moderators in addition to an increased understanding into *how* accurate judgments are formed. In an effort to provide a more nuanced understanding of when and how judgment accuracy is likely to be affected the present study comprises three phases of data collection to explore three specific questions. First, we ask - *when* are we more

[1] In this paper judgement accuracy is assessed via the correlation between a person's personality score (measured as an average of self and informant ratings) and a judgement made about that person's personality by a stranger. We acknowledge that the term 'accuracy' is a loaded term (see also Funder, 1999) and is often measured and defined in terms of self-other agreement (Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Funder & West, 1993; see also Wood & Funder, 2016). As the present study correlates an average of self *and* informant ratings with judgements, the use of 'self-other agreement' would be misleading. We do also acknowledge that the term 'similarity' has been suggested as a more appropriate way by which to describe accuracy scores hence our use of apostrophes and inclusion of this footnote.

likely to make accurate judgments; second, we ask *who* is more likely to be an accurate judge as a function of their judgment perspective (i.e., Interactant or Observer). Finally, we explore the behavioural cues that are used by Interactants versus Observers' when rating another's personality. To the best of our knowledge, the current investigation represents the first study to explore different informational contexts, judgment perspectives and behavioural coding.

1. Theoretical Overview: How do we make Judgments of Others?

Funder (1995, 1999) proposed a model of judgement accuracy known as the Realistic Accuracy Model (RAM). RAM posits that judgment accuracy is a result of a four stage interpersonal and cognitive process (Letzring et al., 2006). Specifically, in order for a successful judgement to be made, a target must display behaviours that are *available* to the judge; those behaviours must be *relevant* to the target's personality; the judge must then *detect* the cues; and finally *utilise* the cues to make an 'accurate' judgment. RAM is interpersonal in the sense that a judge has to either observe or interact with an actual target who makes relevant behaviours available and is cognitive in the sense that a judge has to detect and utilise this target information in order to form an 'accurate' judgment (see Figure 1). Studies exploring RAM have found evidence in favour of what has commonly been referred to as the *trait visibility effect* (Funder & Colvin, 1988; Watson, Hubbard & Weise, 2000), which provides support for RAM. That is, visible traits such as extroversion, or those with frequent and clear behavioural cues, yield better 'accuracy' than less visible traits (e.g., John & Robins, 1993; Watson, 2010). The implication being that if accuracy *is* based on observing the availability of cue relevant behaviour, as implicated in RAM, then there should be an advantage when rating visible traits with numerous cues relative to non-visible traits with less cue availability. In support, the more visible trait of extraversion tends to be the most accurately judged trait in first impression studies (Borkenau & Liebler, 1992; Watson, 1989) and has been shown to correlate significantly with noticeable cues such as physical attractiveness (Albright, Kenny

& Malloy, 1988) and loudness of voice (Borkenau & Liebler, 1992). Conversely, less visible traits such as neuroticism and agreeableness are typically judged with less accuracy (e.g., John & Robins, 1993) and less evidence exists on the behavioural indicators for these traits.

[*Insert Figure 1 about here*]

Utilising this model as a conceptual framework for the current research a number of moderators are relevant to RAM and are the topic of current investigation. Specifically, the literature indicates that there are four moderators of accuracy: judge, target, trait, and information (Funder, 1999, 2001). Research suggests that certain types of people are better judges of another's personality (Human & Biesanz, 2012; Letzring, 2008), that some targets are just easier to judge (Akert & Panter, 1988), and that some traits are simpler to judge than others (Gosling et al., 2002). In the current study we are particularly interested in the final moderator variable – information, as it directly relates to the amount and the quality of the cues that are available to the judge. We propose that if first impression judgments are predominantly driven by the availability of relevant behavioural cues, as posited in RAM, then an increase in the quality and quantity of cues available should increase judgment accuracy.

2. Overview of Current Study

The present three-phase study utilises RAM as a conceptual framework with three major aims: i) to explore the impact of different information contexts on judgement accuracy; ii) to investigate whether judgement perspective (i.e., interactant or observer) influences accuracy and; iii) to identify the behavioural cues that people use when judging another's personality in addition to identifying which of these cues are valid indicators of personality.

Phase 1 employed a 'get to know' task and a 'negotiation' task using a design similar to that implemented previously by Funder and Colvin (1991). Specifically, pairs of

participants completed two 'get to know' interactions with different interaction partners followed by a third 'negotiation' interaction with the same partner as in the second 'get to know' interaction. After each interaction, interacting pairs rendered judgements of each other's personality. Phase 2 employed the data from Phase 1 as stimuli for a group of observers. Specifically, to investigate whether those involved in, or observing, an interaction differ in judgement accuracy, the judgements from observers in phase 2 were compared with the judgments from interactants in phase 1. Finally, phase 3 employed a behavioural coding approach to explore the role of behavioural cues on judgment accuracy for interactants and observers. To the best of our knowledge, no study has compared the impact of judgement perspective on accuracy nor explored the different behavioural cues used by interactans and observers when judging another's personality. Therefore, the present study aims to extend current knowledge of when and how people form accurate judgments when faced with limited information.

3. Phase One: The Role of *Information* on Judgment Accuracy

As implicated in RAM, the information that manifests within a context affects the cues that are *available* and *relevant* to a judge. This information variable is often conceptualised in two ways (Letzring et al., 2006). The first is in terms of *information quantity*, which typically refers to the amount of information available to the judge and thus relates to RAM's availability stage (see Figure 1). For example, evidence showing that friends make more 'accurate' personality judgments than strangers is typically explained by reference to information quantity, since knowing someone for longer means that the judge will have observed more cues from the target's behaviour in more contexts (Jackson, Neill, & Bevan, 1969; Norman & Goldberg, 1966; Paulhus & Bruce, 1992). A similar result comes from varying the breadth of cues available, such as whether the stimulus information available to judges is non-verbal, audio or visual (Reynolds & Gifford, 2001). The second way in which information can differ is in terms of *information*

quality. Information quality relates more closely to RAM's *relevance* stage since different cues can be more or less diagnostic of different aspects of personality (i.e., represent better quality information). For example, Letzring et al. (2006) have shown that judgment accuracy is higher in unstructured 'getting acquainted' contexts versus structured 'debate' situations because behaviour in the latter context tends to be more constrained and thus less diagnostic (i.e., relevant) of personality.

The majority of first impression studies consider information *quantity* (e.g., Blackman, 1995; Funder & Colvin, 1988; Watson & Clark, 1991) and have revealed that more information tends to produce more accurate judgments (Beer & Watson, 2010; Blackman & Funder, 1998; Borkenau & Liebler, 1992; Holleran Mehl, & Levitt, 2009; Letzring et al., 2006). In contrast, much less research has examined how information *quality* shapes judgment accuracy (cf McLarney-Vesotski, Bernieri & Rempala, 2006; Letzring et al., 2006). Importantly, Beer and Brooks note that information quality is largely under examined; thus, the question of what constitutes 'good quality information' when rating another's personality remains an open area of inquiry. The present study, therefore, aims to explore the role of information quality on accuracy. More specifically, it is argued here that one useful way to think about information quality may be in terms of different contexts of interaction which vary in terms of the task and partner.

Research investigating variations in the partner aspect of context have tended to focus on information quantity (Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Paulhus & Bruce, 1992). These studies show that acquainted partners tend to be more accurate at rating each other's personality (e.g., Funder & Colvin, 1991; Norman & Goldberg, 1966; Paulhus & Bruce, 1992). The implication underlying this finding is that the longer we interact with a person the more behavioural information we have available (i.e., information quantity). It is argued here that differences in partner aspects of context are likely to impact not just on information quantity but on information quality. More specifically, interacting with the same

person over time increases not only the amount of behavioural information available but also the quality of information on which to base our judgments (Funder, 1999; Letzring et al., 2006). Put another way, in addition to sharing more information about themselves, people interacting with the same partner multiple times may also be more likely to share information that is relevant about aspects of their personality. However, to the best of our knowledge, the impact of interacting with the same person over time in first impression encounters remains untested.

In addition to theorising about the importance of contextual variations in partner on the trait specific cues available in an interaction, the contextual task may also be important. Specifically, different tasks inherently require different kinds of behaviour and this may impact on the types of information a target reveals. For example, if a target is asked to make a presentation and to answer questions as part of an interview, their behaviour during each of these tasks is likely to differ owing to different demands of the tasks. The majority of research, however, only alludes to the possibility that qualitatively different tasks will shape the relevance of information available (i.e., trait specific cues available) for personality judgments. Although not framed as differences in 'task' *per se*, a number of studies have examined differences between structured and unstructured information contexts (e.g., Carney, Colvin, & Hall, 2007). Structured interaction contexts typically involve some form of constraint either in the topic discussed or in the rules for discussion (Snyder & Ickes, 1985). They therefore give little room for behavioural variation. By contrast, unstructured interaction contexts in which participants 'talk about anything' or 'get to know' each other (e.g., Carney et al., 2007) permit a greater expression of personality relevant information (Borkenau et al., 2004; Funder & Colvin, 1991). Structured contexts such as trivia quizzes (Letzring et al., 2006), debates (Borkenau et al., 2004), and toy building (Blackman, 1996), tend to produce less accurate judgments than are typically found in unstructured contexts and

have led to the conclusion that unstructured interaction contexts render better quality information when rating another's personality. Although the structured-unstructured distinction is useful for understanding the types of contexts that contain good quality information, it represents a broad dichotomy that overlooks trait-specific differences. For example, structured contexts such as 'debates' and 'interviews' which tend to be perceived as more likely to constrain the natural expression of trait relevant behaviour may contain good quality information for specific types of traits. One way to think about such contextual nuances is to consider the mapping between the kinds of cues that are salient to a task and the traits that are likely to relate to those cues (see McLarney-Vesotski et al., 2006). As an example, consider the task of negotiating, which is a task that participants of the present study were asked to complete. Negotiation is a structured task and cues relevant to the traits of openness, conscientiousness and neuroticism are likely to be available. For instance, creating value, also known as *integrative* bargaining, is a process central to negotiation. It involves discovering the other person's interests so that creative solutions can be generated (e.g., Fisher, Ury, & Patton, 1991). This kind of behaviour should elicit more accurate judgments for the trait of openness, since a major facet of this trait concerns imagination and creativity. Second, claiming value, also known as *distributive* bargaining, is the process of taking the resources available and dividing it amongst targets. This behaviour arguably requires a degree of planning about the resources that are most preferable to oneself and others. As conscientiousness is related to social responsibility and planning, as well as to forward thinking (Barrick, Mount, & Strauss, 1993), it would seem plausible to predict that this trait may be judged accurately in negotiation situations. Third, negotiations are affectively charged in nature (Adler, Rosen, & Silverstein, 1998; Kumar, 1997). This may increase the likelihood of trait relevant behaviour for the affectively charged trait of

neuroticism, since people high on this trait tend to be anxious, and highly emotional (Goldberg, 1993).

When thinking about the relations between the cues available in a context and the traits to which such cues relate, the 'get to know' and 'negotiation' tasks may represent useful interaction contexts through which to explore the role of trait specific cues on accuracy. In particular, it would be useful to examine whether negotiation tasks are more diagnostic of particular traits in the ways outlined above. In line with the aforementioned behavioural differences that may occur in the different tasks outlined, the following hypotheses are made:

H1: Judgment accuracy for all Big-5 traits will be more accurate in the *final* negotiation context than the *second* get to know context due to an increase in both the amount of information (i.e., same partner in interactions 2 and 3) and trait specific information (due to different tasks).

H2: Judgment accuracy for the traits of openness, conscientiousness, and neuroticism will be judged more accurately in the negotiation context than in either of the 'get to know' contexts as the quality (i.e., relevance, trait specificity) of information for these traits should be greater.

4. METHOD

4.1 Phase 1 – The role of *information* on judgement accuracy

4.1.1 Participants

53 unacquainted participants received either £8 monetary compensation or 3 course credits for 80 minutes of participation. Participants were recruited via the Universities research participation scheme. Of the 53 participants, a core group of 20[2] Interactants (9

[2] The minimum sample size typically recommended for between subjects' designs is 20 per condition (Simmons, Nelson, & Simonsohn, 2011). Given that within subjects designs typically have more power to detect effects, a sample size of 20 for a within subjects design as part of a series of studies was deemed

males, 11 females) participated in each of the three interaction contexts (Mean age = 21.8, *SD* = 2.3). Self-reported ethnicity of the target participants was 79% White British, 6% Asian, 3% White Irish, 3% Black African and 9% other. The remaining participants comprised 33 Informants who were asked to render judgments about their impression of the Interactants personality. Interactants provided at least one informant rating (*n* = 20), and the response rate for those who received two informant ratings was 65% (*n* = 13). These informant scores were combined with self-ratings and averaged to derive an aggregate measure of target personality (see Wall et al., 2013 for similar approach). The average correlation between self and informant ratings across contexts was *r* = .48 (Range .27 to .69).

4.1.2 Materials

*4.1.3. Personality Measure***.** Participants were asked to rate their personalities using the 50-item International Personality Item Pool measure (IPIP; Goldberg, 2006). The reliability of the 50-item scale was assessed for aggregate personality ratings, and Interactant judgments of each other for each Big-5 trait across contexts. Specifically, for aggregate personality scores (i.e., an aggregate of self- and informant-ratings), separate reliability analyses were conducted on the combined self and informant ratings for each Big-5 trait, resulting in an alpha on either 20 items (i.e., 10 items per trait for self and one informant) or 30 items (10 items per trait for self and two informants) depending on whether an Interactant obtained one or two informant ratings. Reliabilities for the traits of extraversion, agreeableness, conscientiousness, neuroticism, and openness were .93, .76, .79, .92, and .89 for the aggregate personality ratings and were .89, .87, .89, .84, and .89 for the judgment ratings.

4.1.4 Procedure

---

sufficient for present purposes (see also Wilson VanVoorhis & Morgan, 2007). Importantly, replication is needed to reduce false positives (Fraley & Vazire, 2014).

Participants signed up to a two-part study advertised as 'Interpersonal Perception between Strangers'. On arrival to the laboratory, pairs of Interactants were allocated to separate rooms and introduced to the study. As the present study comprised a within subjects design, they were instructed that they would be asked to participate in a total of three interactions over a two week period and informed that the interactions would vary in terms of what they would be asked to do (i.e., their 'interaction task'), and in terms of who they would interact with (i.e., their 'partner'). The order of tasks was consistent for all participants. To go beyond some of the limitations with self-reported personality (Alicke, 1985; Kwan, John, Kenny, Bond, & Robins, 2004; Paulhus & Trapnell, 2008), participants were asked to provide personality ratings from up to two knowledgeable informants (i.e., people who knew them well) so that aggregate personality ratings could be obtained. As noted in the material section, informant scores were combined with self-ratings and averaged to derive an aggregate measure of target personality (see Wall et al., 2013 for similar approach).

4.1.5. *Interaction one* (New Partner). In this task, participants were separately taken to a video lab, sat opposite their fellow Interactant, and were left to interact without the experimenter present. They were instructed to talk about "anything they like". The lab contained unobtrusive video cameras on which their interaction was recorded (participants were aware of this). After ten minutes the experimenter returned and separately escorted each participant back to their original location so that no further interaction could take place, and they were asked to fill out an IPIP questionnaire about their impression of their interaction partner's personality (IPIP-other rating). Participants were then debriefed and scheduled to return approximately one week later for interactions two and three.

4.1.6. *Interaction two* (New Partner, Same Task). Participants were randomly allocated a partner that was different to that of their first interaction. The rest of the experiment proceeded as interaction one.

4.1.7. *Interaction three* (Same Partner, Different Task). This interaction took place after Interactants had finished rating their interaction partner from interaction two. The task was a modification of the employment contract negotiation task employed by Olekalns and Smith (2000). In this task, one person was randomly allocated the role of 'employer' and the other person the role of 'employee'. They then role-played a negotiation over a job contract for the position of 'Graduate Recruitment Consultant', a position chosen because of its relevance to students looking for graduate employment[3].

After negotiating for ten minutes participants were individually taken to their separate rooms and asked to render a judgment of their interaction partner's personality, again using the IPIP-other rating. As they had already rated their Interaction partner in interaction two, it was important to ensure that participants did not feel that the purpose of this interaction was to change their previous judgment. Thus, participants were assured that responses may be the same *or* different and that there is no right or wrong answer. In an effort to avoid judgment ratings being affected by completion of the self-report all participants were asked to complete the self-oriented IPIP as a rating of their own personality *after* they had completed all interactions and all IPIP-other ratings, and were then debriefed.

5. Results and Discussion

5.1. Analytic Strategy. Interactants (i.e., person being judged) personality was measured using a composite of self and informant ratings (see Letzring, Wells, & Funder, 2006; Wall, Taylor, & Campbell, 2016 for similar approach). Judgment 'accuracy' was

[3] Although Interactants were instructed to try to reach an agreement, they were also informed that they did not have to reach an agreement; thus, if an aspect was difficult to agree on then they could move on to another issue. This was designed to allow for the expression of individual differences in behaviour as it stops people agreeing if they do not want to. Indeed, other than the standard instructions and pay-off schedule, participants were free to use any other information to ensure realism and flow of task

measured by correlating judgments of Interactants' personality with Interactant's composite personality score. When selecting an approach to assess accuracy it is important to be cognisant of the different measurement approaches (see Cronbach & Gleser, 1959; Wood & Furr, 2016). One way by which to assess accuracy is by correlating responses across trait items for each target-judge pair, known as profile correlations (i.e., rendering as many scores as judges). It has been suggested that this method depends strongly on the extent to which each variable reflects "normative" information (Furr, 2008; Wood & Furr, 2016) and factors like the scoring of the items. A second method is the item approach which correlates scores across persons for each trait or item and was the approach adopted here. That is, item level correlations were calculated and then aggregated resulting in as many accuracy scores as there were items[4]. Our decision to use the item approach was due to our focus on trait level predictions in addition to interpretational issues with profile scores. It is important to note that there are limitations associated with each approach (see Cronbach & Gleser, 1959; Funder, 1999), and the error which likely stems from each approach arguably cancels itself out across the three interaction conditions (see Letzring, 2008 for similar rationale on issues of non-independence).

Means and SD's were computed for judgements accuracy per trait across the three interaction conditions (see Table 1). In order of magnitude, extraversion was the most accurately judged trait across contexts (M = .16, SD = .18), followed by neuroticism (M = .17, SD = .14), agreeableness (M = .03, SD = .12), conscientiousness (M = .01, SD = .12), and finally openness (M = -.19, SD = .18) across contexts. This pattern of trait 'accuracy' is in line with previous research which tends to report that extraversion is typically the most

[4] We acknowledge that researchers in this area tend to aggregate items prior to correlating; however, as we compare accuracy scores across conditions, as opposed to within condition, our analyses would not have been possible using summed scale scores as this would have produced only one score per trait preventing univariate comparisons. Currently, we have ten scores per trait that can be compared across contexts to address our research questions.

accurately judged trait with openness often being found to be one of the more difficult traits to judge (Albright, Kenny, & Malloy, 1988; Funder & Colvin, 1988).

[*Insert Table 1 about here*]

Recall that the first hypothesis predicted that judgment accuracy for all Big-5 traits would be more accurate in the final negotiation context than the *second* get to know context due to an increase in both the amount of information and trait specific information. The hypothesis was supported for the trait of conscientiousness, $t(9) = -2.24$, $p = .04$ whereby accuracy was higher in the final negotiation task ($M = .07$; $SD = .15$) than the second 'get to know' task ($M = -.01$; $SD = .13$). Interestingly, the reverse effect was found for neuroticism as there was a significant difference between these two contexts, $t(9) = 4.03$, $p = .01$ whereby accuracy was higher in the second 'get to know' task ($M = .37$; $SD = .19$) than the final negotiation task ($M = .17$; $SD = .23$). Counter to hypotheses, analyses found that accuracy was not significantly different in the second 'get to know' task than the final negotiation task for the traits of extraversion, $t(9) < 1$, $p = .91$, agreeableness, $t(9) = -1.52$, $p = .16$ .05 or openness, $t(9) < 1$, $p = .84$.

The second hypothesis that the negotiation task would elicit increased judgment accuracy for the traits of conscientiousness, neuroticism and openness when compared to the *first* 'get to know' task was examined using a series of paired samples *t*-tests for each Big-5 trait with item correlation scores as the Dependent Variable. The hypothesis was not supported for conscientiousness, $t(9)$[5] $= -1.62$, $p = .14$, neuroticism, $t(9) = -1.64$, $p = .14$, nor

[5] Although there were 20 participants per condition the degrees of freedom are 9 for each comparison as an item analysis was performed with ten IPIP items per trait (see Wall et al., 2013 for similar analysis). Although the present within subjects design meets the minimum sample size required for a between subjects design a post hoc power analysis was performed for each comparison and the range of power reached was in an acceptable range of .70-.90 (see Cohen, 1988 for specific recommendations; see also Fraley & Vazire, 2014 for debate on issue of post hoc power). Moreover, the scores are based on an aggregate of self and informant ratings which aims to enhance validity.

openness, $t(9) = < 1$, $p = 99$. Thus, qualitative differences in the tasks examined did not impact on the accuracy of Big-5 judgments.

It is useful to further investigate the finding that neuroticism was rated more accurately in the second get to know task than the third negotiation task as this was counter to predictions ($p = .01$). This finding suggests that participants may elicit important behavioural differences in the second get to know task; thus, it would seem useful to also compare accuracy for the first and second get to know tasks to further understand these findings. A series of paired samples *t*-tests comparing accuracy scores in the first two get to know interactions revealed a significant effect for the trait of neuroticism only, which bolsters the findings reported above for this trait. Specifically, accuracy was higher in the second 'get to know task' ($M = .37$; $SD = .19$) than the first, ($M = -.04$; $SD = .29$); $t(9) = -3.23$, $p = .01$. The present phase suggests that familiarity with a task may shape judgment accuracy for the less interpersonal and typically difficult to judge trait of neuroticism. Although this finding is interesting and highlights a differential pattern of trait accuracy across interaction contexts it remains unclear whether this enhanced accuracy for neuroticism in the second 'get to know task' relates to more cues being revealed (i.e. enhanced availability in RAM) and/or better detection of cues (i.e., enhanced detection/utilization in RAM). Findings from Funder and Colvin's (1991) study support our suggestion that it may be due to people feeling more relaxed which may indicate enhanced detection/utilisation. The findings for conscientiousness revealed a different pattern to that of neuroticism and results were in line with our prediction that more information would enhance judgement accuracy. When considering *how* we form judgments of others these findings tentatively suggest that additional cues may facilitate better judgments. Moreover, when considering the implications of these findings for RAM the data suggest that different stages of RAM may be more influential when judging specific traits. Specifically, increased detection/utilisation appears to

be useful when judging neuroticism, which is in line with definitions of this trait as difficult to detect through actual behaviour. In contrast, accuracy for conscientiousness appears to be more relevant to the availability and relevance stages of RAM as accuracy for this trait increased when more information was available.

6. Phase 2: The Role of *Perspective Type* on Accuracy

To further understand judgment accuracy in first impression encounters, it is useful to consider another moderator of accuracy – the *judge*. Personality psychologists have dedicated years of research to examining the personality characteristics of a 'good judge' (e.g., Allport, 1937; Taft, 1955; Vogt & Colvin, 2003). However, as noted by Letzring (2005) no variables have repeatedly emerged as reliable correlates of accuracy (Davis & Kraus, 1997). Although it is tempting to conclude that no reliable correlates exist, it is argued here that a focus on the task demands of the judge may be a useful avenue of investigation. Specifically, it is important to go beyond individual characteristics of judges such as their level of extraversion (Ambady, Hallahan, & Rosenthal, 1995) and focus more on the informational perspective of judges as this may impact on the salience of cues on which judges rely. Indeed, judges observing an interaction may be more, or less, accurate when rating a target's personality than a judge directly involved in conversation as the level of engagement in an interaction may impact on the processing style judges engage in. Proximate support for this assertion comes from a study by Giordano, George, Marett and Keane (2011) who note that Observer judgments can suffer as they cannot follow up with questions and are removed from the surrounding context that often contains subtle nuances. Such studies, however, have focused largely on judgments of deception (e.g., Bonito, Burgoon, Ramirez, & Dunbar, 2000) and it is unknown whether these differences will remain when rating another's personality.

Moreover, differences have also been found in the way that Interactants and Observers process information. For example, Interactants tend to rate others more favorably (Burgoon & Newton, 1991), and have also been found to remember and interpret cues differently (Stafford, Waldron, & Infield, 1989). Pronin, Fleming and Steffel (2008) also found that Targets and Observers perceive self-disclosure differently in terms of the value within them. Indeed, a fundamental difference is that Interactants are almost always engaging in multiple activities at once (Gilbert, Jones, & Pelham, 1987); thus, it is reasonable to expect that personality judgments made by Interactants may suffer relative to that of Observers. Taken together, the weight of the existing evidence indicates that Observers may have a preferential advantage when rating an Interactants personality, and leads to the following hypothesis:

H3: Observers will be more accurate than Interactants when rating personality across context.

6.1 Participants

Eighty participants (males = 33, females = 47) with a mean age of 20.1 years, (*SD* = 2.01) were recruited from the same University as phase 1 via the Universities research participation scheme. Observers also received either £8 monetary compensation or 3 course credits for 80 minutes of their time. The interaction data from phase 1 served as stimuli for the Observers.

6.2 Procedure

On arrival to the laboratory, participants were fully informed that the study was interested in 'Interpersonal Perception between Strangers'. Consistent with the practice of others in the field (e.g., Back et al., 2010; Vazire, 2007), Observers worked in groups of four so that a mean Observer rating could be computed. Specifically, each Observer was asked to watch three interactions recorded in a previous study (i.e., phase 1) so that they could rate

three different targets. This allowed each target to be rated by four Observers and also enabled meaningful comparisons to Interactants' judgments. Specifically, Observers viewed two 'get to know' interactions and one negotiation interaction ensuring that Interactants and Observers each rated three different targets in total.

Moreover, as a single recording depicted a pair of Interactants on the screen, Observers were explicitly requested to code only the behaviour of *one* of the two people on the video. Once Observers knew which Interactant they had to observe, they were instructed to focus exclusively on that person, even when the other Interactant was speaking. Observers were asked to rate their impression of the target's personality using the IPIP employed in phase 1. On completion of their rating of three Interactants, Observers were thanked and debriefed.

7. Results & Discussion

The extent of convergence across all four Observers (i.e., consensus for the judgment data) was assessed using intraclass correlations ICC (2, *k*) (Shrout & Fleiss, 1979), computed using a one-way random effects model (see also Back et al., 2010). In order of magnitude, extraversion had an average ICC of .97 across all Observers, followed by conscientiousness, .96, neuroticism, .95, agreeableness, .94 and openness, .91. These high ICC's justify averaging across Observers, as is common practice in the field (e.g., Back et al., 2010). These average scores produce reliability estimates of .93 for extraversion, .89 for agreeableness, .92 for conscientiousness, .89 for neuroticism and .79 for openness.

a. Who is a better judge of another's personality?

When examining accuracy correlations the order of magnitude for Observers was similar to that obtained for Interactants in phase 1. Specifically, extraversion was the most accurately judged trait ($M_{\text{Observers}}$ = .32, *SD* = .24 vs. $M_{\text{Interactants}}$ = .16), followed by

neuroticism (*M* = .16, *SD* = .21 vs. $M_{\text{Interactants}}$ = .17), conscientiousness (*M* = .13, *SD* = .13 vs. $M_{\text{Interactants}}$ = .01) agreeableness (*M* = .03, SD = .17 vs. $M_{\text{Interactants}}$ = .03), and finally openness (*M* = .03, *SD* = .11 vs. $M_{\text{Interactants}}$ = -.19). The finding obtained in phase 1 that extraversion, in order of magnitude, was the most accurately judged trait and openness the least accurate, was replicated for Observers.

The third hypothesis that Observers would be more accurate than Interactants was tested using a series of independent *t*-tests. This hypothesis was partially confirmed as significant differences were found between Interactants and Observers when rating the less interpersonal traits of conscientiousness across contexts, *t*(9) = -3.52, *p* = .01, *r*[6] = .76 and openness *t*(9) = -3.24, *p* = .02, *r* = .74. Specifically, Observers (*M* = .13, *SD* =.04) were more accurate at rating conscientiousness than Interactants (*M* = .01, *SD* = .04) and were also more accurate at rating openness (*M* =.03, *SD* = .04) than Interactants (*M* = -.19, *SD* = .06). Significant differences were not found for neuroticism, *p* = .95, extraversion, *t*(9) = -1.67, *p* = .11, or agreeableness, *t*(9) = < 1, *p* = .98.

To follow up on the finding that Observers were more accurate than Interactants when rating conscientiousness and openness and determine whether this interacts with context a 3 (Context: first 'get to know' task, second 'get to know task', third 'negotiation' task) by 2 (Judgment Perspective: Interactant, Observer) repeated measures ANOVA was conducted for these two traits. In support of the above comparisons, a main effect of perspective-type was found for conscientiousness *F*(1, 18) = 5.15, *p* = .04, η_p^2 = .22 and openness *F*(1, 18) = 11.07, *p* = .001, η_p^2 = .38 whereby Observers were more accurate than Interactants when rating these traits. These findings held across context as no interaction effect was found for

[6] Effect size *r* was computed using an effect size calculator (https://www.easycalculation.com/statistics/effect-size-t-test.php)

conscientiousness, $F(2,36) = 1.46$, $p = .255$, $\eta_p^2 = .08$ or openness, $F(2,36) = 1.03$, $p = .29$, $\eta_p^2 = .10$.

Recall that a major aim of this phase was to develop the findings regarding the role of information on accuracy in phase 1. Specifically, the findings from phase 1 that judgments for neuroticism were more accurately rated in the second 'get to know' task than the final negotiation task were attributed to increased familiarity which may have enhanced interacants *detection* and/or increased the amount of cues *available*. Thus, if the present effect was due to Interactants revealing more relevant cues to their personality then Observers should also be able to detect these cues. As Observers were able to rate conscientiousness and openness more accurately than Interactants *across all contexts* this offers tentative support that adopting an Observers perspective enhances detection and utilisation. We acknowledge that this claim is tentative and warrants further exploration using a behavioural coding approach.

The pattern of trait accuracy for Interactants vs. Observers is interesting. More specifically, the finding that Observers were more accurate when rating the less interpersonal traits of openness and conscientiousness (Gill & Oberlander, 2003; Gosling et al., 2002) suggest differences in *how* these judgment perspectives utilise cues: Do Observers notice different cues and/or utilise cues differently to Interactants? One useful way in which to explore these questions is to examine the behavioural cues revealed in the different interactions alongside an exploration of how such cues differentially relate to Observers' *and* Interactants' judgments. Accordingly, the final phase explored this issue using a behavioural coding approach.

8. Phase 3: The Role of *Behavioural Cues* on Personality Judgement Accuracy

Recall that Funder's (1995, 1999) model of judgment accuracy posits that in order for people to form 'accurate' judgments about another's personality *relevant* cues to that

persons' personality need to be *available* and such cues need to be *detected* and correctly interpreted (i.e., *utilised*). When theorising about how people form judgments this is a useful framework through which to further our understanding of judgement accuracy. By implication we might conjecture that to understand how we form accurate judgments we need to explore the ways in which personality is revealed through *behaviour* in order to determine what specific cues are actually available and whether or not judges are detecting and utilising these cues.

A number of studies have examined the link between behaviour and personality. Existing research has focused largely on the trait of extraversion (e.g., Scherer, 1978, 1986), which has found that extraverts tend to 'ask more questions', are more likely to 'tell jokes', and tend to 'initiate more conversations' than introverts (Argyle, Martin, & Crosland, 1989; Thirne, 1987). Gosling and colleagues (2002) found that the ways in which a person adorns their bedroom is indicative of their level of openness. More recently, evidence suggests that a great deal about our personality is contained within email addresses (Back, Schmukle, & Egloff, 2008), static appearance cues (Vazire, Naumann, Rentfrow, & Gosling, 2008) and a person's clothing (Naumann et al., 2008), and that some of these cues can be reliably inferred by judges. These studies are substantive as they require behavioural coding and in so doing encourage researchers to examine 'real' behaviour in 'real contexts' (Baumeister, Vohs, & Funder, 2007). Building on the behavioural coding approach to understanding personality-behaviour links, the present study explores the behavioural cues that interactants *and* observers appear to use when making their judgments.

9. Method

9.1 Participants

Forty participants (males = 16, females = 24), all undergraduate students with a mean age of 23.9 years, (*SD* = 2.56), signed up to partake in this behavioural coding study and were paid £8 for 80 minutes of their time. Participants were again recruited via the Universities research participation scheme and the study was advertised as being interested in 'Personality and Behaviour'.

9.2 Materials

9.2.1. Behaviour stimuli. The interaction data from phase 1 served as stimuli for a team of coders. Thus, the data set comprised a total of 60 interactions, each lasting approximately ten minutes.

9.2.2 Behaviour Questionnaire. A 26-item questionnaire for measuring behaviour was developed for the study. Specifically, in an effort to sample behaviour relevant to each Big-5 trait, 12 items were selected from the 100-item IPIP as this measure contains questions that directly ask about behavior (see Table 2). An item was selected if it represented a behaviour that could feasibly be coded from viewing an interaction. For example, the IPIP item 'starts conversation' measures a facet of extraversion and represents a behaviour that should be accessible to those observing the interaction. Similarly, an example of agreeable behaviour that could also be coded from viewing an interaction was "inquires about others'; an item for conscientiousness was 'appears neatly dressed'; an item for neuroticism was 'appears anxious', and an item for openness was 'catches onto things quickly'.

The behavioural items for the more evaluative traits, such as neuroticism and openness, were harder to select as they largely concern feelings and, by definition, relate to behaviour that is less visible. For example, an item measuring agreeableness that was selected from the IPIP was '*feels little concern for others*.' This item arguably relates to a feeling that would need to be inferred and thus could not be assessed 'objectively' by coders. Consistent with Funder and Colvin (1991), who adapted the Q-sort in a similar fashion, additional items were

created on the basis of previous research findings (e.g., Berry & Hansen, 2000). For example, the additional behavioural items of '*fidgets*', '*self-touches*', '*open body language*', '*uses hands when talking*', '*smiles'* and '*nods head*' were added. In support of the inclusion of these behaviours, previous research has reported significant correlations between 'open' body positions and personality (Berry & Hansen, 2000). Also, 'fidgets', 'nods head' and 'anxious' have also been employed in Ambady and Rosenthal's (1998) study and 'smiles' were related to a person's level of extraversion in Kenny's (1992) study. It should be noted, that behaviours such as 'smiles' and 'nods head' are typically coded in concrete ways but were coded here in general terms in an effort to capture mid-level, and more stylistic behaviours. Thus, no coder counted how many times a person smiled or nodded their heads, but someone would receive a high score for 'smiles' if they engaged in the behaviour frequently.

The decision to adapt a more macro approach to coding behaviour, considering behaviour style rather than the instances of a behaviour's occurrence is supported by Funder and Colving (1991) who note that this is often a more valid way to code behaviour. The focus on more general, stylistic behaviours is arguably more relevant to the present focus on the Big-5 as these traits concern a person's average level of behaviour in 'real life' contexts when forming judgements of targets, since judges are unlikely to count the number of times the target smiles but may be more likely to employ a general feel for how a person behaves. Trained coders have shown themselves to be proficient at reliably coding behaviour in this way (Cairns & Green, 1979; See also Funder & Colvin, 1991).

All items were measured on a 7-point scale that ranged from (1) extremely inaccurate to (7) extremely accurate. For example, when coding whether a person 'fidgets', participants were informed that 'extremely accurate' refers to 'an extremely accurate representation of that person's behaviour' (i.e., they fidget a lot) whereas 'extremely inaccurate' refers to 'an extremely inaccurate representation of that person's behaviour' (i.e., they never fidget).

9.3 Procedure

Groups of four participants (i.e., behavioural coders) were asked to watch a total of six clips each containing a pair of Interactants engaging in audio and video recorded conversation. They were instructed that the study was interested in how people behave during an initial encounter with a stranger, and that their task was to code the behaviour of only one of the two people in the video. They were told to focus on that person even when the Interactant was listening to their partner. It was emphasised that they must try to remain objective throughout the coding procedure and refrain from making any judgments about the Interactant whom they had been asked to code. Thus, coders were explicitly instructed to code observable behaviour and not behaviour that they expected to see or inferred their target would do on other occasions. No participant studied the same Interactant's behaviour in more than one interaction condition. This ensured that analyses concerning overlap between cues valid and cues utilised (H4) were not confounded with the coding procedure (see also Funder & Colvin, 1991).

The experimenter was present throughout the behaviour coding to ensure that each coder independently rated Interactants' behaviour without influence from any other coders in the room. Participants received no additional training instruction, which is consistent with previous research showing that naive raters can produce reliable ratings without in-depth training (e.g., Rosenthal, 1987; Rosenthal, Blanck, & Vannicelli, 1984).

10. Results & Discussion

As each coder rated a total of 6 Interactants, the scores are non-independent; thus, to err on the side of caution, a second measure of agreement is needed that takes this issue into consideration. As recommended by Shrout and Fleiss (1979) the ICC (2, *k*) agreement statistic was employed. The average agreement amongst coders was equivalent to that found above

ICC_{mean} = .87 (Range .64 to .94). Means for each behaviour and Cronbach alphas were computed for each interaction (see Table 2).

[*Insert Table 2 about here*]

10.1 Analytic Strategy for Personality-Behaviour Links

The Brunswik (1956) model explicates how a researcher can examine the process of personality judgments as articulated in RAM (see Figure 2). Specifically, *availability* may be assessed through the coding of behaviour itself. In other words, if a team of research assistants can reliably code a target's behaviours, then those behaviours must in some sense be available (see Figure 2, A). The notion of *relevance* concerns the cues that are valid indicators of personality (i.e., cues valid). In Brunswik's model this is measured by the cues that correlate with personality (see Figure 2, B). RAM's third stage of *detection* refers to the cues that judges seem to employ in their judgments of a target, which may or may not represent valid cues to the target's personality. In Brunswik's model this is measured by the cues that correlate with judgments formed (see Figure 2, C), and is referred to as 'cues utilised'. To fully integrate RAM and Brunswik's model, the overlap (measured via correlations) between cues valid and cues utilised should lead to judgment accuracy, and thus successful utilisation (see Figure 2, D).

[*Insert Figure 2 about here*]

To examine more directly the relation between cues utilised and cues valid, the fourth hypothesis predicted that the most accurately judged traits (in order of magnitude) would demonstrate the highest levels of overlap between cues utilised and cues valid. This issue required an additional set of correlations. Specifically, as recommended by Steiger (1980; and employed by Gosling et al., 2002 and Funder & Sneed, 1993), inter-correlations were

performed between the cue-utilisation correlations and the cue-validity correlations for Interactants *and* Observers using Fishers *r*-to-*z* transformation (and were transformed back to *r* for presentation). The correlations were computed for each Big-5 trait. In other words, the overlap between cues utilised and cues valid were examined to determine what Brunswik called 'achievement' (i.e., the achievement of accuracy) (see Table 3).

[*Insert Table 3 about here]*

Recall that, in order of magnitude the traits that were most accurately judged for Interactants were extraversion, *(*r_{mean} = .17), neuroticism (r_{mean} = .16), agreeableness (r_{mean} = .03), conscientiousness (r_{mean} = .01) and finally openness (r_{mean} = -.19). A similar pattern of accuracy was found for Observers whereby extraversion was the most accurately judged (r_{mean} = .32), followed by neuroticism, (r_{mean} = .16), conscientiousness, (r_{mean} = .13), agreeableness (r_{mean} = .03) and finally openness to experience was also the least accurately judged trait for Observers (r_{mean} = .03). As shown in Table 3, the hypothesis that a significant level of overlap would be found between cue utilisation and cue validity correlations was partially confirmed. Specifically, the hypothesis was confirmed for both Interactants, r = .50, p < .05, and Observers, r = .64, p < .01 for the trait of extraversion suggesting that valid cues were used when rating this most accurately rated trait.

Recall in phase 2 that when Observers' ratings were compared with Interactants' ratings, Observers were more accurate when rating the less interpersonal traits of conscientiousness and openness. Interestingly, in terms of the intercorrelations between cue validity (i.e., correlations between cue and self-reported personality) and Observers cue utilisation correlations (i.e., correlations between Observers' judgments formed and cue) a significant level of overlap was found for the trait of conscientiousness, r = .51, p < .05 but not for the trait of openness. However, it should be noted that although the correlation for

openness was low for Observers, it was positive, whereas the overlap between cue utilisation and cue validity for Interactants was negative for openness, $r = -.66$, $p < .01$.

To further understand the differences in accuracy between Interactants and Observers accuracy, it would seem useful to compare, at a descriptive level, the intercorrelations between cue utilisation and cue validity for Interactants and Observers. As shown in Tables 4 to 6, intercorrelations for extraversion indicated that Interactants appeared to detect and utilise valid interpersonal type behaviours such as '*starts conversation*', $r = .61$, '*quick to understand* ', $r = .42$ and '*laughs*', $r = .47$ whereas Observers appeared to accurately detect and utilise cues related to assertive type behaviours such as '*takes charge*', $r = .50$, and '*starts conversation*', $r = .48$ as well as interpersonal type behaviours such as '*appears at ease*', $r = .43$ and '*smiles*', $r = .54$. Similarly, for agreeableness, Interactants detected and utilised cues such as '*agrees with partner*', $r = .60$, '*moans*', $r = -.55$ and '*smiles*', $r = .45$, whereas Observers detected and utilised cues such as '*uses hand when talking*', $r = -.54$ and '*inquires about others*', $r = .58$. Interestingly, cue utilisation correlations for conscientiousness appear to be very different for Interactants and Observers. That is, Interactants had no significant cue validity correlations (with the exception of 'takes conversation to a higher level) whereas Observers correctly detected the valid cue of '*nods head*', $r = .52$, '*avoids eye contact*', $r = -.51$ and 'pays *attention to detail*', $r = .59$ when rating the trait of conscientiousness. For neuroticism, Interactants appeared to utilise the cue of '*fidgets*', $r = -.56$ whereas Observers detected the cue of '*anxious*', $r = -.47$ and '*quick to understand*, '$r = .42$ (see Tables 4 to 6).

Given the lower accuracy reported for openness in both phases (relative to the other traits) the finding that this trait produced the smallest number of cue utilisation correlations (see underlined values in Tables 4 to 6) those of Borkenau and Liebler (1995) and Funder and Sneed (1993) who also found that judges seem to employ the wrong cues when judging this

trait. This supports the lower accuracy found for this trait in general and assertions that this trait is difficult to detect through overt behaviour. Interestingly, this finding suggests that the lower accuracy found for openness may be explained by judges detecting the wrong cues as opposed to a lack of valid cues *per se*. The trait of conscientiousness stood out as showing the biggest discrepancy between Interactants and Observer for cues detected. Specifically, out of the cues examined, Interactants did not appear to use any of the valid cues available when rating their impression of their partner‘s personality (see Tables 4 and 5). Interestingly, this may partially explain the low accuracy found for Interactants relative to Observers and again supports RAM's claim that accurate judgments are based on detecting valid cues. As behavioural coders were not asked to count how many times a person engaged in behaviours (i.e., number of eyebrow twitches) the salience of a behaviour may have affected the perceived frequency (e.g., a broad smile may have led to higher ratings than 2-3 minor smirks)[7]. Although this is possible, we believe that our findings are in line with numerous researchers in this field (e.g., Borkenau & Liebler, 1995; Funder & Colvin, 1991; Funder & Sneed, 1993). Also, as we focus on relative accuracy across contexts this potential error variance should cancel itself out across the three contexts examined.

When viewed together, the findings suggest that although the number of cues available for conscientiousness were low, Observers were able to utilise these cues more successfully than Interactants across contexts. In terms of openness, findings suggest that Observers found judging this trait difficult (r = .05) but Interactants appeared to find judging this trait even more difficult (r = -.66; See Tables 4 to 6). It will be useful for future research to determine if the present findings can be replicated and conduct research that directly asks participants what cues they used in their judgments.

[7] We would like to thank an anonymous reviewer for pointing this out. It will be interesting for future research to directly explore this possibility.

11. General Discussion

The present study sought to develop our understanding of *how* accurate personality judgments are formed. To explore this question phase 1 examined *when* we are more likely to be accurate (Get to know vs. Negotiation contexts), phase 2 examined *who* is more likely to render an accurate judgment (Interactant vs. Observer) and phase 3 explored this question more directly and examined *how* we form judgments (Behavioural Coding) in addition to how Observers and Interactants may differ in this process. Results indicated that differences in tasks examined (i.e., get to know vs. negotiation) did not impact on accuracy but that differences in partner did for the trait of conscientiousness. Our findings also revealed important differences in trait specific accuracy in terms of whether a judge was an Interactant or an Observer whereby Observers were more accurate when rating the less interpersonal traits of conscientiousness and openness.

The findings in phase 1 provided some support for the differential role of the amount of cues available (but not trait specific cues) on accuracy by showing that accuracy for conscientiousness was higher in the negotiation interaction where participants interacted with the same partner, however, accuracy for neuroticism was higher when participants engaged in a second 'get to know' interaction with a different partner. A plausible explanation is that quantity of information can be useful for conscientiousness as participants arguably had more cues available over two different interactions with the same partner. In contrast, familiarity with the task appears to be useful when rating traits such as neuroticism. Tables 4 to 6 showed that more cues were utilised in interaction two than in interaction one suggesting that familiarity may increase detection. Although this task did not find that manipulating the amount of trait specific cues available through a negotiation task increased judgment accuracy it will be interesting for future researchers to explore other tasks that may be more

effective, building on the work of Gosling et al., (2002). To provide further support for this claim future researchers could ask Interactants and Observers to report how they felt within the interaction in terms of 'comfortableness', cognitive load, and motivation to judge accurately, among others.

Phases two and three went some way to developing our understanding of the role of the amount of information (information quantity) as opposed to the amount of trait specific information (information quality) on accuracy as they revealed that the types of cues that Interactants' and Observers' focus on differed by trait and in terms of the type of cue that was relevant. Specifically, results showed that Observers were more accurate when rating the less interpersonal traits of conscientiousness and openness whereas Interactants were more accurate when rating the "visible" trait of extroversion. Our finding that the inherently more "visible" trait of extroversion was rated more accurately by Interactants, relative to other traits, corroborates research in this area (Albright, Kenny, & Malloy, 1988; Funder & Colvin, 1988). Our finding that Observers were less accurate when rating this trait is an interesting finding given the high visibility often attributed to this trait and suggests that the salience or visibility of a trait may relate to perspective type.

Further support for the link between perspective type and saliency of cue comes from the cue level analysis which revealed that the type of cue that Observers detected differed to Interactants in interesting ways. Specifically, cues related to "turn-taking" and "interpersonal type" behaviours (e.g., laughs, nods head) were found for Interactants whereas assertive type behaviours (e.g., takes charge) were noticed by Observers – (i.e., who stood out vs. how someone responds). The differential salience of cues has previously been discussed (Wall et al., 2013) and our findings suggest that different judgement perspectives may attend to and utilise different types of cues, which is in line with RAM (Funder 1995, 1999). That is, even when a range of relevant information is available an accurate judgement will not result if

judges do not notice or use it in their assessments of others. In support of this explanation, Ames and Bianchi (2008) found an "agreeableness asymmetry" whereby supervisees but not supervisors were more inclined to notice whether their supervisor was warm and trustworthy whereas supervisors we more inclined to notice cues relevant to their supervisees reliability. Although substantive their study looked at perceptions (i.e., cue utilisation) and not cue validity; thus, the present findings develop and extend this line of theorising and suggest that people seem to attend to different cues which can impact on the accuracy of personality judgements. These findings are interesting and warrant further investigation and replication. One promising avenue for future investigation would be to examine why Observers detect cues relevant to different traits. Indeed, other studies suggest that although the process of accuracy may be data driven (i.e., bottom up) it is largely based on pre-existing knowledge. In support, a study by Biesanz, West, and Millevoi (2007) found that when judges were less acquainted with a target, and thus presumably had less individuating information, judgements were based more on generalised (i.e., stereotypical, top-down) knowledge as opposed to knowledge about a specific person.

When theorising about *how* accurate judgements are formed, our finding that Observers were accurate for certain traits supports Funder's (1995) suggestion that moderators of accuracy interact. Funder commented that a judge by trait interaction should be termed *expertise*. That is, he argues that such a finding may arise due to differential knowledge of trait behaviour relations or self-serving differences in judges' conceptualisations of traits (see Funder, 1995, p. 663). As the sample was randomly selected and comprised undergraduate students it is unlikely that knowledge of trait behaviour-relations differed systematically across conditions. The more parsimonious explanation arguably relates to self-serving differences between Interactants and Observers (i.e., motivational factors). Indeed the results of phase 3 are in line with this possibility as the same

number of valid cues were available to Observers and Interactants but the type of cue they appeared to utilise differed in terms of the interpersonal nature of the trait. Put another way, interpersonal traits might be less salient to Observers as knowing how sociable someone is likely to be is arguably less relevant to an Observer who will not interact with this person (see also Wall, Taylor & Campbell, 2016).

Although the major aim of this three-phase study was to increase our understanding of how we make accurate personality judgements across information contexts it is important to consider the contextual variations in cues. At a descriptive level, it can be gleaned from Tables 4, 5 and 6 that the cues that were valid indicators of personality in one context were not always valid indicators of personality in another context. Although it could be argued that this is evidence against traits as consistent patterns of behaviour our findings are actually in line with Funder and Colvin's (1991) substantive study. Specifically, they showed that behaviour can and does change across context but that people still demonstrate some consistency in behaviour. It is clear from the tables that some behaviours were linked to personality traits in more than one context whereas others were not. This therefore supports Funder and Colvin's conclusion that behavioural consistency and change across context can occur.

11.1 Limitations and Future Directions

A number of limitations need to be borne in mind when considering the findings of the present study. We acknowledge that our exploration into the availability of trait specific cues in phase 1 could have been more systematic in terms of directly disentangling the effects of amount of information and trait specific information in addition to addressing order effects. However, we believe our exploratory three phase study goes some way to addressing the differential role that the amount of information and trait specific information has on

judgement accuracy and encourage further research in this underexplored area. Indeed, it should be acknowledged that it is incredibly difficult to directly investigate information as the construct itself has been deemed 'an ambitious realm that is largely undefined (Beer & Brooks, 2011, p. 176; cf Letzring et al., 2008; Beer & Watson, 2010) but we feel that our investigation of the availability of the amount of trait specific cues provides a good starting point.

Another potential criticism that can be levied against the studies presented relates to the small effect sizes. Importantly, when comparing these effect sizes to other research some studies have reported Big-5 accuracy correlations that range from .05 to .13 (Back et al., 2008; see also Naumann et al., 2009). As judgements were compared across context it is relative accuracy (i.e., accuracy across context) as opposed to absolute accuracy (i.e., mean level) that is of interest here. As recommended by Cohen it is important to interpret effect sizes in context (Glass *et al.* 1981, p104). The results in the present study examined judgements made between unacquainted individuals (i.e., strangers); therefore, given the limited information that participants had about each other and the time constraints placed on the interaction, these findings are interesting and warrant further exploration.

It is possible that our study did not have enough power to detect significant effects given the small to moderate sample size. However, the comparisons across conditions using a repeated measures design reduces this issue. Additionally, a number of significant findings emerged that were in line with trait definitions and converged with literature surrounding valid cues to personality (e.g., Costa & McCrae, 1992; Gosling et al., 2002) and accuracy research (Albright, Kenny, & Malloy, 1988; Funder & Colvin, 1988). A strength of the current study lies in the focus on 'real' people in 'real' contexts (Funder, 1999) and the exploration of 'behaviour' using the same methodological approach throughout.

Nevertheless, as observers in phase 2 judged the personality of the moderate sample from phase 1 it is crucial that the present study is replicated.

It is possible that, due to the sample comprising undergraduates who typically engage in "getting acquainted" activities, the familiarity of the task may have impacted upon findings. Although this is possible and warrants replication in a more heterogeneous sample, it is important to note that the present aim was to understand how people form *initial* personality judgements. Findings should not be generalised beyond first impression encounters, which Funder (1999) also acknowledges as the social and cognitive process involved in judging the personality of a well acquainted person may differ.

In conclusion, our findings suggest that when faced with more information (increase in the amount of cues as well as trait specific cues; phase 1) judgement accuracy increases for conscientiousness only. Observers were found to be more accurate than Interactants when rating the less interpersonal traits of conscientiousness and openness (phase 2) and the role of cues appear to be important (phase 3). Critically, our findings suggest that cues differentially shape judgement accuracy processes depending on whether we are engaged in an interaction or observing an interaction and this, we believe, is an important part of theory development that requires further research attention. Put another way: "When Looking at the same interaction different information perspectives may see something different".

References

Adler, R. S., Rosen, B., & Silverstein, E. M. (1998). Emotions in negotiation: How to manage fear and anger. *Negotiation journal*, *14*(2), 161-179.

Akert, R. M., & Panter, A. T. (1988). Extraversion and the ability to decode nonverbal communication. *Personality and Individual Differences*, *9*(6), 965-972.

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of personality and social psychology*, *55*(3), 387.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. Journal of Personality and Social Psychology, *49(6),* 1621.

Allport, G. W. (1937). *Personality* (pp. 173-181). New York: Holt.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, *69*,(3), 518.

Ambady, N., & Rosenthal, R. (1998). Nonverbal communication. *Encyclopedia of mental health*, *2*, 775-782.

Ames, D. R. & Bianchi, E. C. (2008). The agreeableness asymmetry in first impressions: Perceivers' impulse to (mis)judge agreeableness and how it is moderated by power. *Personality and Social Psychology Bulletin, 34(12),* 1719-1736.

Back, M. D., & Nestler, S., (2016). Accuracy of Judging Personality. In J. A. Hall, M., Schmid Mast, & T. V.West (Eds.). *The social psychology of perceiving others accurately*. Cambridge University Press

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

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological science*, *21*(3), 372-374.

Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology*, *78*(5), 715.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396-403.

Beer, A., & Brooks, C. (2011). Information quality in personality judgment: The value of personal disclosure. *Journal of Research in Personality*, *45*(2), 175-185.

Beer, A., & Watson, D. (2010). The effects of information and exposure on self-other agreement. *Journal of Research in Personality*, *44*(1), 38-45.

Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin*, *20*(4), 367-378.

Berry, D. S., & Hansen, J. S. (2000). Personality, nonverbal behavior, and interaction quality in female dyads. *Personality and Social Psychology Bulletin*, *26*(3), 278-292.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social psychology*, *92*, (1), 119.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology*, *34*(2), 164-181.

Bonito, J., Burgoon, J. K., Ramirez Jr, A., & Dunbar, N. E. (2000). Participation and decision-making: The role of interactivity in communication processes and outcomes. *Manuscript submitted for publication*.

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

Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63(1),* 1-25.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of personality and social psychology*, *86*(4), 599.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press.

Burgoon, J. K., & Newton, D. A. (1991). Applying a social meaning model to relational message interpretations of conversational involvement: Comparing Observer and participant perspectives. *Southern Journal of Communication*, *56*(2), 96-113.

Cairns, R. & Green, J. (1979). How to assess personality and social patterns: Observations or ratings? In Cairns (Ed.), *The analysis of social interactions*: Methods, issues and illustrations, p209-226.

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

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. *Hillsdale, NJ*, 20-26.

Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and individual differences*, *13*(6), 653-665.

Cronbach, L. J., & Gleser, G. C. (1959). Interpretation of reliability and validity coefficients: Remarks on a paper by Lord.

Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. *Empathic accuracy*, 144-168.doi:10.1016/j.jrp.2005.08.007

Fisher, R., Ury, W., & Patton, B. (1987). *Getting to yes*: Negotiating agreement without giving in. (2nd ed.). New York, NY: Penguin books.

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: evaluating the quality of empirical journals with respect to sample size and statistical power. *PloS one*, *9*(10), e109019.

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychological review*, *102*(4), 652.

Funder, D. C. (1999). *Personality judgment: a realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C. (2001). Accuracy in personality judgment: Research and theory concerning an obvious question.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of personality and social psychology*, *55*(1), 149.

Funder, D. C. & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality, 60(5),* 773.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: an ecological approach to judgmental accuracy. *Journal of personality and social psychology*, *64*(3), 479.

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of personality*, *76*(5), 1267-1316.

Gill, A., & Oberlander, J. (2003). Perception of e-mail personality at zero-acquaintance: Extraversion takes care of itself; neuroticism is a worry. In *Proceedings of the 25th annual conference of the cognitive science society* (pp. 456-461). Hillsdale, NJ: Erlbaum.

Giordano, G., George, J., Marett, K., & Keane, B. (2011). Reviewers and the Detection of Deceptive Information in Recorded Interviews. *Journal of Applied Social Psychology*, *41*(2), 252-269.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: personality judgments based on offices and bedrooms. *Journal of personality and social psychology*, *82*(3), 379.

Gilbert, D. T., Jones, E. E., & Pelham, B. W. (1987). Influence and inference: What the active perceiver overlooks. *Journal of Personality and Social Psychology*, *52*(5), 861.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, *48*(1), 26.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public domain personality measures. *Journal of Research in personality*, *40*(1), 84-96.

Holleran, S. E., Mehl, M. R., & Levitt, S. (2009). Eavesdropping on social life: The accuracy of stranger ratings of daily behavior from thin slices of natural conversations. *Journal of Research in Personality*, *43*(4), 660-672.

Human, L. J., & Biesanz, J. C. (2012). Accuracy and assumed similarity in first impressions of personality: Differing associations at different levels of analysis. *Journal of Research in Personality*, *46*(1), 106-110.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of personality*, *61*(4), 521-551.

Kumar, R. (1997). The role of affect in negotiations: An integrative overview. *The Journal of Applied Behavioral Science*, *33*(1), 84-100.Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. Guilford Press.

Kwan, V. S., John, O.P., Kenny, D. A., Bond, M. H. & Robins, R. W. (2004). Reconptualizing indivual differences in self-enhancement bias: An interpersonal approach. *Psychological Review, 111(1),* 94-110. Doi: 10.1037/0033-295X.111.1.94

Letzring, T. D. (2005). Personality and behavioural correlates of accurate personality judgment. *Unpublished doctoral dissertation,* University of California.

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of research in personality*, *42*(4), 914-932.

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

McLarney-Vesotski, A. R., Bernieri, F., & Rempala, D. (2006). Personality perception: A developmental study. *Journal of Research in Personality*, *40*(5), 652-674.

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

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, *4*(6), 681.

Olekalns, M., & Smith, P. L. (2000). Understanding optimal outcomes. *Human Communication Research, 26*(4), 527-557.org/10.1016/j.jrp.2007.01.004

Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology*, *63*(5), 816.

Pronin, E., Fleming, J. J., & Steffel, M. (2008). Value revelations: disclosure is in the eye of the beholder. *Journal of Personality and Social Psychology*, *95*(4), 795.

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, *8*(4), 467-487

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (3rd ed.). Reading, MA: Addison Wesley.

Stafford, L., Waldron, V. R., & Infield, L. L. (1989). Actor-Observer differences in conversational memory. *Human Communication Research, 15*(4), 590-611.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological bulletin*, *87*(2), 245.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, *52*(1), 1.

Vazire, S. Naumann, L.P., Rentfrow, J. J. & Gosling, S. D. (2008). Portrait of a narcissist: Manifestation of narcissism in physical appearance. Journal of Research in Personality, 42(6), 1439-1447.

Vernon, P. E. (1933). Some characteristics of the good judge of personality. *The Journal of Social Psychology*, *4*(1), 42-57.

Vogt, D. S., & Randall Colvin, C. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, *71*(2), 267-295.

Wall, H. J., Taylor, P. J., & Campbell, C. (2016). Getting the balance right? A Mismatch in Interaction demands between Target and Judge Impacts on Judgement Accuracy for Some Traits but not Others. *Personality and Individual Differences, 88,* 66-72.

Wall, H. J., Taylor, P. J., Dixon, J., Conchie, S. M., & Ellis, D. A. (2013). Rich contexts do not always enrich the accuracy of personality judgments. *Journal of experimental social psychology*, *49*(6), 1190-1195.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self report. *Journal of Personality and Social Psychology*, *57*(1), 120.

Watson, D. (2000). *Mood and temperament*. New York: Guilford Press

Watson, D., & Clark, L. A. (1991). Self-versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology*, *60*(6), 927.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of personality and social psychology*, *78*(3), 546.

Wood, D., & Furr, R. M. (2016). The correlates of similarity estimates are often misleadingly positive: The nature and scope of the problem, and some solutions. *Personality and Social Psychology Review*, *20*(2), 79-99.

Table 1. Interactants Cue Utilisation and Cue Validity Correlations as a Function of Interactions in Phase 1 (continues on next page)

	Extraversion			Agreeableness			Conscientiousness			Neuroticism			Openness		
	Interaction														
Behavioural Cue	Int1	Int2	Int3	Int1	Int2	Int3	Int1	Int2	Int3	Int1	Int2	Int3	Int1	Int2	Int3
Smiles a lot			.58**	.45*		.43ǂ				-.44ǂ					
Uses difficult words			.45*										.41ǂ		
Neatly dressed	-.44ǂ	-.46*													
Laughs a lot	.47*		.52*		.42ǂ		.43ǂ				.51*				
Inquires about Others		.52*		.55*									-.43ǂ		
Full of ideas	-.62**			.44ǂ			-.44ǂ			-.56**					
Fidgets		-.51*		-.44ǂ	-.48ǂ -.40ǂ										
Takes conversation to a higher level							.52*			.54*			.48ǂ	.42ǂ	
Shows they are listening	.43ǂ	.50*		.43ǂ	.45*										
Pays attention to detail		.57**	.57**	.51*									-.55*		
Agrees with partner		-.56**		.60**		.47*				.52*	.55*	.55*		-.54*	-.54*
Self touches	-.48*	.50*					-.46*								
Open body language		.57**				.42ǂ							.44ǂ		
Interrupts													.55*	.51*	
Takes charge		.44*													.53*
Anxious		-.51*													
Acts at ease											-.41ǂ				
Appears relaxed		.51*													
Avoid eye contact		-.44ǂ													
Talks about self									-.52*						
Moans		-.56**		-.55*	-.45*										
Does not probe			.44ǂ												
Starts conversation		.61**			-.44ǂ	.54*									
Quick to understand	.42ǂ										-.43ǂ				

Note. * $p < .05$. ǂ $p < .10$, two-tailed. Values underlined represent cues valid. All other values represent cue utilisation correlations. I1, I2 and I3 refer to whether the interaction was interaction one, two or three.

Table 2. Mean Behaviours for Each Interaction Situation

Behaviour	Interaction 1	Interaction 2	Interaction 3	Cronbach's Alpha
1. Starts conversation[1]	4.70	5.33	4.36	.83
2. Quick to understand things[5]	5.56	5.45	5.10	.78
3. Smiles a lot	4.68	4.55	4.24	.89
4. Uses difficult words [5]	3.74	3.60	3.84	.81
5. Appears neatly dressed	5.08	4.65	4.79	.78
6. Laughs a lot	4.08	4.09	3.65	.86
7. Takes charge[1]	4.46	4.78	4.49	.86
8. Inquires about others[2]	4.53	4.55	3.43	.87
9. Talks about self	4.70	4.53	4.44	.88
10. Full of ideas[5]	4.34	4.35	4.55	.82
11. Fidgets	4.16	4.21	3.58	.89
12. Carries conversation to high level[5]	4.39	4.56	3.88	.86
13. Nods head	4.44	4.44	3.59	.88
14. Appears anxious[4]	3.40	3.41	3.40	.89
15. Appears at ease[4]	4.63	4.71	4.38	.87
16. Waits for others to lead the way	3.56	3.39	3.89	.88
17. Moans about things[4]	3.40	3.40	3.54	.87
18. Shows they are really listening	5.44	5.38	4.95	.90
19. Appears relaxed most of the time[4]	4.58	4.81	4.49	.90
20. Avoids eye contact	3.36	3.11	3.75	.89
21. Pays attention to details[3]	5.03	5.13	4.84	.81
22. Agrees with partner	5.06	5.48	4.29	.75
23. Self-touches	4.35	4.24	3.63	.89
24. Open body language	4.69	4.39	4.33	.90
25. Uses hands when talking	4.51	4.78	4.20	.86
26. Interrupts	3.03	3.10	3.26	.92

Note. Numbers in superscript denote items taken and adapted from the IPIP-Big5 (see Method section, page 21). [1] relates to behaviour items relevant to the trait of extroversion with an internal consistency of .93, [2] relates to behaviour items relevant to the trait of agreeableness (internal consistency cannot be computed as only one item taken form IPIP). [3] relates to behaviour items relevant to the trait of conscientiousness with an internal consistency of (internal consistency cannot be computed as only one item taken from IPIP. [4] relates to behaviour items relevant to the trait of neuroticism with an internal consistency of 67. [5] relates to behaviour items relevant to the trait of openness to experience with an internal consistency of .70.

Table 3. Overlap between Behavioural Cue Utilisation and Cue Validity Correlations for Interactants and Observers

Big-5 trait	**Interactants**	**Observers**
Extraversion	.50**	.64**
Agreeableness	-.09	-.06
Conscientiousness	-.23	.51**
Neuroticism	-.26	-.19
Openness	-.66**	.05

Note. * $p < .05$. ** $p < .001$, two-tailed.

Table 4. Interactants and Observants Cue Utilisation and Cue Validity Correlations as a Function of Interactions in Interaction 1 in Phase 1

	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Behavioural Cue	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt
Smiles a lot		.44‡	.45*	.49*			-.44‡			.55*
Uses difficult words									.41‡	
Neatly dressed	-.44‡							-.47*		
Laughs a lot	.47*	.51*			.43‡					
Inquires about others			.55*						-.43‡	
Full of ideas	-.62**		.44‡		-.44‡		-.56**			
Fidgets			-.44‡							
Takes conversation to a higher level					.52*		.54*		.48‡	
Shows they are listening	.43‡		.43‡					.44*		
Pays attention to detail			.51*						-.55*	
Agrees with partner			.60**				.52*			
Self touches	-.48*				-.46*					
Open body language									.44‡	
Interrupts									.55*	
Takes charge		.50*								
Appears Anxious						-.47*		-.47*		
Appears at ease		.43‡					-.41‡	.55*		
Appears relaxed										
Avoid eye contact						-.44*				
Talks about self										
Moans			-.55*							
Starts conversation										
Quick to understand	.42‡						-.43‡	.42‡		
Nods head		.46*								.43*
Waits to be led										-.58**
Uses hands when talking				-.54*						-.50*

Note. * $p < .05$. ‡ $p < .10$, two-tailed. Values underlined represent cues valid. All other values represent cue utilisation correlations. I1, I2 and I3 refer to whether the interaction was interaction one, two or three.

Table 5. Interactants and Observants Cue Utilisation and Cue Validity Correlations as a Function of Interactions in Interaction 2 in Phase 1

	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Behavioural Cue	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt
Smiles a lot										
Uses difficult words										
Neatly dressed	-.46*									
Laughs a lot	.52*						.51*	.51*		
Inquires about others			.55*	.58**						
Full of ideas			.44ǂ							
Fidgets	-.51*		-.44ǂ							
Takes conversation to a higher level									.42ǂ	
Shows they are listening	.50*		.43ǂ							
Pays attention to detail	.57**		.51*							
Agrees with partner	-.56**		.60**	.49*			.55*	.55*	-.54*	
Self touches	.50*									
Open body language	.57**									
Interrupts									.51*	
Takes charge	.44*					.42ǂ				.46**
Anxious	-.51*							-.44ǂ		
Appears at ease							-.41ǂ			
Appears relaxed	.51*									
Avoid eye contact	-.44ǂ									
Talks about self										
Moans	-.56**		-.55*							
Starts conversation	.61**			.59**						.57**
Quick to understand							-.43ǂ	.55**		
Nods head						.52*				
Waits to be led										
Uses hands when talking										

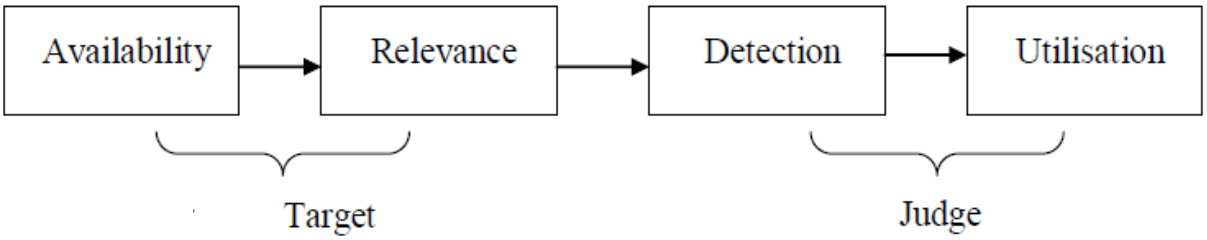
Note. * $p < .05$. ǂ $p < .10$, two-tailed. Values underlined represent cues valid. All other values represent cue utilisation correlations. I1, I2 and I3 refer to whether the interaction was interaction one, two or three.

Table 6. Interactants and Observants Cue Utilisation and Cue Validity Correlations as a Function of Interactions in Interaction 3 in Phase 1

	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Behavioural Cue	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt	Int'nt	Obsv'nt
Smiles a lot	.58**	.54*	.43ǂ					.79**		
Uses difficult words	.45*									
Neatly dressed										
Laughs a lot	.52*	.44ǂ						.48*		
Inquires about others							-.43ǂ			
Full of ideas					-.56**	.59*				.52*
Fidgets										
Takes conversation to a higher level					.54*		.48ǂ			
Shows they are listening						.59*				
Pays attention to detail	.57**					.59**				.42ǂ
Agrees with partner		.52*	.47*	.50*	.52*		-.55*			
Self touches						-.42ǂ				
Open body language			.42ǂ	.57**		.60**	.44ǂ			.61**
Interrupts		.46*				-.42ǂ	.55*			
Takes charge								-.49*		
Anxious		.52*				.52*				.58**
Appears at ease										
Appears relaxed						.71**				.61**
Avoid eye contact						-.51*				-.42ǂ
Talks about self										
Moans						-.74**				
Starts conversation		.48*	.54*							.51*
Quick to understand						.44ǂ				
Nods head						.68**		.48*		
Waits to be led		-.48*								
Uses hands when talking										

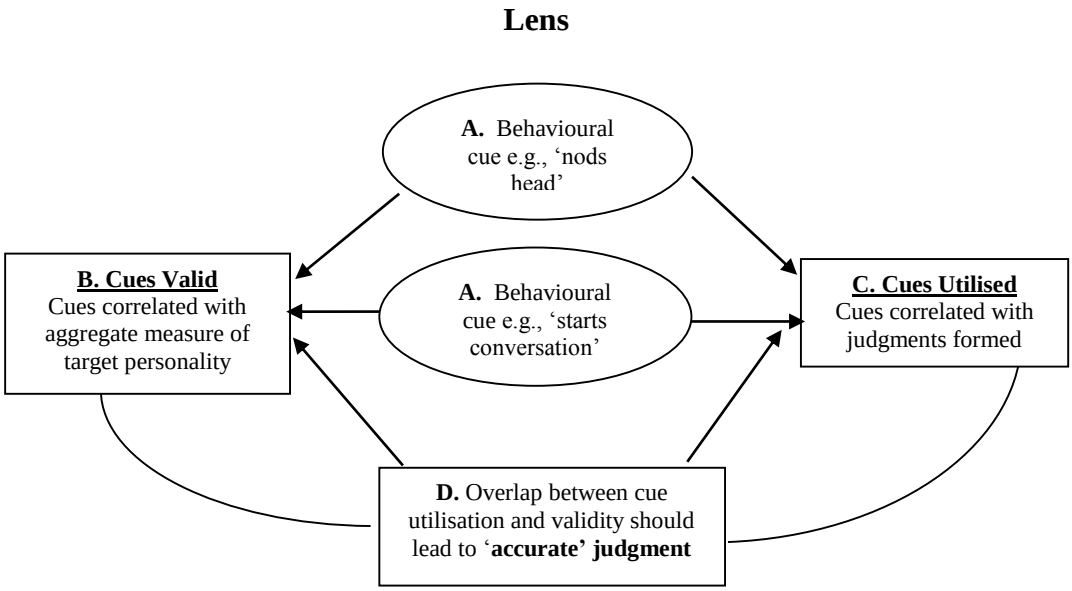
Note. * $p < .05$. ǂ $p < .10$, two-tailed. Values underlined represent cues valid. All other values represent cue utilisation correlations. I1, I2 and I3 refer to whether the interaction was interaction one, two or three.

Figure 1. Visual Representation of Funder's (1995) Realistic Accuracy Model



Note. Figure adapted from Funder (1999) and Letzring (2005).

Figure 2. Brunswik Lens' Model of Inferential Behavior



Note**. A** refers to the behavioral cues in the environment (RAMS's *availability* stage); **B** refers to cues valid (RAM's *relevance* stage); **C** refers to cues utilized (RAM's *detection* stage); and **D** refers to the overlap between cues utilized and cues valid (RAMs' *utilization* stage)