From photograph to face-to-face:

Brief interactions change person and personality judgments

Liam Paul Satchell*[a]

*Corresponding author.

Can be contacted at: liam.satchell@winchester.ac.uk, University of Winchester, Sparkford Road, Winchester, Hampshire, SO22 4NR

[a] Department of Psychology, University of Winchester

Author note: The data for this research can be found on the Open Science Framework, here: https://osf.io/bx8zv/

Acknowledgements: This study is made possible thanks to the contributions of data collectors and coders working for course credit. Thanks go to: Ella Strutt, Alex Brown, Sumaya Shaho, Shola Radford, Braden Shrimpton, Akshaiya Jothrirajan, Tasnim Ahmed, Bolu Olowoyo, Bradley Nicholls, Idara Itauma, Riya Tiwari, Stefunmi Olatunji. Further thanks go to Eglantine Julle-Danière, Charlotte Hudson, Oliver Waddup and Harry Mayes for comments on a draft of this paper and Dr Nina Tupper and Jackson Tupper (TupperTwins) for their assistance with visualisation of results.

Highlights

- Person(ality) judgments before (using photos) and after interaction are studied
- Person(ality) judgments changed after a brief (<5 minute) interaction
- After interaction; some Big Five and psychopathic personalities were more accurate
- After interaction; judgments were more pro-social and less negative
- There was some continuity in judgments made before and after interaction

Abstract

Research into 'first impressions' frequently uses photographs of faces as representations of unknown others. This is the methodological solution to concerns about standardisation, sample size and experimental control. However, there is little work investigating the robustness of these first impressions through to first interactions. A sample of 97 pairs of stranger participants (N = 194) completed personality measures (Big Five and Triarchic Model of Psychopathy), made personality and social judgments of a photograph of the face of their to-be partner, then engaged in five minutes (maximum) of unstructured interaction with their partner and then made their judgments again. The behaviour of the participants in the interaction was coded using 76 criteria by 13 trained coders. Results showed that, judgments made before and after interaction were correlated, but different. Personality judgment accuracy at Time 1 was poor overall but improved at Time 2 participants for Neuroticism, Extraversion and psychopathic Boldness. Coded behaviours related to 'engagement' were those that influenced the person judgments the most, and these were related to Extraversion, Agreeableness and Boldness of participants. Overall, the results of this study show that first impressions of personality change from photographs to face-to-face interaction. Person judgment research should be aware of the extent to which judgments of photographs relate to first interaction.

Keywords: First impressions; Person judgment; Photographs; Ecological Validity

From photograph to face-to-face:

Brief interactions change person and personality judgments

Introduction

"The theories and experiments described here all refer to an essentially passive onlooker, who sees someone do something (or sees two people do something) and then makes a judgment about it. He (this is the generic passive he) doesn't do *anything – doesn't mix it up with the folks he's watching, never tests his judgments in action or in interaction."* Ulric Neisser (1980, p.603, emphasis and brackets from original).

The above Neisser quote is extracted from a commentary the leading psychologist wrote after attending a symposium on 'social knowledge' in 1980. Nearly 40 years ago, Neisser wrote about how social perception research was reliant on studying how individuals passively observe representations of social information more than how individuals behave in social settings. Neisser's critique that the social perception research does not involve social interaction (Neisser's 'mixing it up with folks') still stands for most cases of person judgment and social perception research today. The field is populated with many studies where 'first impressions' of another person are tested using photographs of faces or simple video excerpts. The reasons for this passive, standardised, approach to social perception are justified in a contemporary social psychology that is facing challenges with the replicability, reliability and statistical power of findings. However, it could be argued that we do not truly understand the nature of *in vivo* first impressions, when photograph-based research dominates the field of person judgment research. Revolution in addressing replicability and reproducibility could be improving the quality of research that does not well represent the analogous everyday context. The current study focuses on comparing standard person judgment paradigms to a realistic context. Here, the stability of person and personality judgments based on the typical experimental task (a photograph of an unknown person) is compared with judgments based on a first interaction.

There is a large body of social psychology that is interested in how we perceive other people. These perceptions of others can lead to bias or prejudice which could, in turn, lead to later problematic interaction. However, much of the evidence for first impressions is based on notably

asocial stimuli; static images of faces. Many studies have tested perceptions of others on important social criteria such as gender (such as Marlowe, Schneider, & Nelson, 1996; Sutherland, Young, Mootz, & Oldmeadow, 2014), race (such as Gonzalez, Steele, & Baron, 2017; Wilson, Hugenberg, & Rule, 2017) and sexuality (such as Rule & Ambady, 2008; Rule, Bjornsdottir, Tshkay, & Ambady, 2016) using only photographs of faces. Emotion, which is functionally dynamic in everyday life, is also often investigated using photographs of faces (such as Masuda, et al., 2008; Olszanowski, et al., 2015; Sabatinelli et al., 2011). Even studies that highlight the importance of environment and context of face perception also use photographs of faces (such as Brambilla, Biella & Freeman, 2018; Caroll & Russell, 1996). It is somewhat surprising, given the widespread use of this methodology, that little work has investigated the robustness of photograph perceptions after interacting with a person.

The reduction of a person to a standardised presentations has four main benefits. A fixed set of stimuli for presentation enhances: *i*) replicability of the study procedure, *ii*) efficiency of the study (thus improving opportunities for better statistical power and *N*), *iii*) a decomposition of key features of interest (for example isolating the face) and *iv*) experimental control over varying components of the targets (removing glasses, clothing hair style). The first two benefits are easy to justify given the current issues with replicability in psychology (Open Science Collaboration, 2012) and the need for larger sample sizes (small effects are typical in social and personality psychology; Richard, Bond & Stokes-Zoota, 2003). The latter two benefits pose concerns about the ecological validity and representativeness of research designs (Brunswick, 1956, see Araújo, Davids & Passos, 2007). It is common to present highly 'controlled' faces that lack everyday variability in facial expression, hair presence, glasses, facial jewellery or makeup (i.e. Carré, Morrisey, Mondloch & McCormick, 2010; Little, 2014; Talamas, Mavor & Perrett, 2016; Willis & Todorov, 2006). However, these are critical social signals that individuals use to signal themselves to the world. For example, 'neutral' facial expressions are not normal in social interaction. Smiling or dynamically acknowledging an other person is an important prosocial process. As such, neutral expressions can carry a negative valence and even infants consider unmoving faces unpleasant (see Adamson & Frick, 2003). Makeup (Jones & Kramer, 2016; Jones, Russell & Ward, 2015), hairstyle (Kim & Lee, 2011) and beards (Dixson &

Vasey, 2012) are all known to affect various social perceptions, and removing them from facial stimuli takes away how individuals would choose to present themselves. In fact, despite the large areas of research on makeup, hairstyle, beards, facial morphology (Carré, & McCormick, 2008; Costa, Lio, Gomez, & Sirigu, 2017; Deska & Hugenberg, 2018; Deska, Lloyd, & Hugenberg, 2018; Geniole, Denson, Dixson, Carré, & McCormick, 2015; Hashelhuhn, Ormiston, & Wong, 2015; Hehman, Leitner, & Gaertner, 2013; Ormiston, Wong, & Haselhuhn, 2017), and many other facial variants, we know very little about social perception when these factors are all present together – as would be the case on first meeting. It may well be the case that certain effects (such as pupil dilation as a social cue; van Breen, De Dreu & Kret, 2018) in research do not survive contact with other effects (such as gesture; Hillman, Vrij & Mann, 2012), with these 'bubbles' of research bursting when brought together. This 'bubble-ism' in the literature has limited everyday application of perception research. Researchers would struggle to predict, for example, the perceived attractiveness of an individual presented in a holistic format. We could show evidence that perceived attractiveness is highest for female targets with a lower face width-to-height ratio (Geniole et al., 2015), wearing cosmetics (Jones et al., 2015) and a 'long wave hairstyle' (Kim & Lee, 2011) *separately*, but the effect of the gestalt combination of these facets is unknown. Further, we do not know the weighted implication of cosmetics with a higher face width-to-height ratio or any other opposing effects (for the numerous other facial variants) in the typical limited stimuli bubbles of research. In sum, through distilling a stimulus to a simpler, standardised, presentation, person perception research has developed theory for specific features of a person, but has limited understanding of the acquisition of everyday social knowledge (as noted by Neisser, 1980).

Impoverishing stimuli to introduce experimental control presents opportunities for naturally occurring variability to inform our research. It is possible to let variance occur and account for variability with post-hoc coding. This coded information could be used in analytic techniques, such as Brunswick's (1952) concept of a lens model, to understand the influence of stimulus variability on topics of interest (for a review, see Karelaia & Hogarth, 2008). A lens model analyses the relationship between a stimulus' *i*) judged qualities, *ii*) observable properties and *iii*) their criterion quality. This

procedure explores Brunswick's definition of ecological validity details the relationship between perceived and 'actual' qualities of the world (for a review see Araújo, et al., 2007). With this approach, concerns about difference in participants' appearance, such as 'unusual or unconventional appearance' can be accounted for with post hoc coding (Funder, Furr, & Colvin, 2000).

Social psychology regularly codes for behaviour (as can personality psychology, see Furr, 2009), so it is not novel to produce measures of stimulus presentation variability in naturalistic settings. In the personality judgment literature, the Realistic Accuracy Model (RAM; Funder 1999; 2012) highlights the importance of decomposing for behaviour to best understand judgment accuracy. The RAM has four stages and emphasises that judgments of traits are accurate when behaviours *relevant* to personality traits are *available* for judges to *detect* and *utilize* for accurate judgment. The behaviours relevant to personality judgment can be investigated by looking at the relationship between self-reported trait and actions in the world, for example by using coding schemes such as the Riverside Q-Sort (Funder et al., 2000). A neat hairstyle could be a relevant behaviour for the judging of Conscientiousness if a relationship between coded hairstyle tidiness and self-reported target Conscientiousness was established. Availability refers to the opportunistic manifestation of the relevant behaviours. For example, the neat hairstyle might be covered by a hat (or selectively removed in the presentation of a cropped stimulus photo). If hairstyle is indicative of Conscientiousness but the hairstyle is missing from a presentation, then accuracy at judging a person's traits may be lower than everyday life. Detection in the RAM is the observation of the available and relevant information. If an individual's neat hairstyle is relevant and available for observation, a judge's attention may be distracted or obstructed and thus does not detect the cue. Therefore, even when a target has readily available relevant behaviours, accuracy can be poor due to inattention. Finally, a judge must utilize detected behaviour in order to form a judgment. Bias, prejudice, context and other various personal and environmental influences may lead a judge to interpret the useful information in an incorrect manner. For example, the 'neat' hairstyle may have a cultural meaning, where it is assumed by the judge that "people with that hairstyle are unreliable", compromising broader Conscientiousness judgment accuracy. The RAM is a good framework for understanding personality judgment and

highlights the problems with the limitation of available stimulus person information with the intention of experimental control.

In the present investigation, participants are given the most opportunity to act as they wish in the interaction phase. This creates the natural opportunity for the sample to generate their own available relevant information for detection and utilization. A participant's personality may lead them to generate more observable behaviour indicative of their traits. Labelled the 'expressivity halo' (Bernieri, Gillis, Davis & Grahe, 1996), it is the case that extraverted behaviour is the easiest behaviour to detect in first impressions research (Albright, Kenny & Malloy, 1998; Bernieri, Zuckerman, Koestner & Rosenthal, 1994; Funder & Colvin, 1988) as extraverts display the most relevant behavioural information. The opportunity for many traits to occur at their most natural is present in this current study due to *i*) moving from photograph presentations to interactive dialogue and *ii*) unstructuring the interaction time.

As other authors have noted, it is increasingly common that the first visual inspection of another person is by photograph (Vazire & Gosling, 2004; Naumann, Vazire, Rentfrow & Gosling, 2009). Photographs of faces are becoming a normal aspect of online social and business networking websites and it is important to know how these presentations affect first impressions into first interactions. Personality judgments from photographs of faces have been shown to be accurate predictors of traits (Borkenau, Brecke, Möttig & Paelecke, 2009; Gordon & Platek, 2009; Gosling, Augustine, Vazire, Holtzman & Gaddis, 2011; Naumann, et al., 2009). Face perception research has suggested that from a variety of possible judgments that can be made of faces, the perceived qualities of a face can be effectively summarised on two (principal component analysis-derived) domains; valence and dominance (Oostergof & Todorov, 2008; Walker & Vetter, 2009; Wang, Hahn, DeBruine & Jones, 2016; *c.f.* Sutherland et al., 2018; Wang et al., 2018). These domains of judgment are considered to reflect the potential risks and rewards of interacting with the presented person.

It is right to focus research on faces too, as the faces attract social attention and are highly informative (Bruce & Young, 2012). However, social perception can be informed by the body (Morrison, Wang, Hahn, Jones & DeBruine, 2017) and interactivity (Albright, et al., 1998; Brown &

Bernieri, 2017; Biesanz et al., 2011; Paulhus & Reynolds, 1995; Satchell & Pearson, 2017). Indeed, social interaction is best informed by active perception, where a perceiver can probe another person for more information about their disposition (see Good, 2007). With this in mind, it is surprising that there is limited study of the stability of first impressions of target people from photographs to first impressions of target people from interaction.

There has been important recent research on how increased interactivity can benefit person judgment accuracy. Brown and Bernieri (2017) recently studied accurate first impressions of Big Five personality traits over a period of 10 weeks. On first meeting, a cohort of students judged the personality traits of five to seven unacquainted individuals in a round-robin paradigm (see Brown & Bernieri, 2017, Figure 1). This 'zero acquaintance' judgment occurred before opportunities for interaction. There was evidence of accurate judgments of Extraversion and Openness from this limited contact. Over the following week participants had dedicated interactions with individuals from their round-robin group. These lasted five minutes and were seated "getting-to-know-you" conversation (see Brown & Bernieri, 2017, Figure 2). After these short discussions personality judgment accuracy improved across all domains, with Extraversion judgments being the most accurate and there was acceptable detection of Neuroticism, Agreeableness and Conscientiousness. Over the following weeks there were dedicated activities to provide more information for the members of the round-robin group. A final personality judgment session, held 10 weeks after the zero acquaintance session, revealed further improvements to accuracy across the Big Five traits, with Extraversion and Agreeableness judgments being notably accurate. This study is a highly naturalistic investigation of how individuals come to know each other better and demonstrates the improving accuracy of judging personality traits over time. The current study builds on Brown and Bernieri's work and investigates the particular behaviours and aspects of self-presentation that may be facilitating personality judgment accuracy. Further, the participants' first interaction will be unstructured and organic, with an open space to interact however they wish, in contrast to previous work.

Previously, research has shown that increased contact with another person reduces the uncertainty in opinion about an other person (Sunnafrank, 1986). Whether this opinion becomes more favourable or not after interaction is debated. Norton, Frost and Ariely (2007) investigated perceived liking of another person before and after a date. They found that liking of the individual on the date decreased as they learned more information (reported on adjective scales) about them. This result was questioned by Reis, Maniaci, Caprariello, Eastwick and Finkel (2011a) who observed *in vivo* interactions between two previously unacquainted individuals, and found that familiarity increased liking. These two papers were debated in comment articles (Norton, Frost & Ariely, 2011; Reis, Maniaci, Caprariello, Eastwick, & Finkel, 2011b) before the two labs came together to suggest that the methodological definitions of 'familiarity' matters (Finkel et al., 2015). The present study takes the form of a 'get acquainted interaction', in Finkel et al.'s terminology. These paradigms tend to improve liking of another in interaction (Reis et al., 2011a; for an overview, see Finkel et al., 2015). Liking is not a main focus of the current investigation (it is included for general interest), however it could be the case that general positivity about the targets of the judgments could be affected in similar ways. This could compromise personality judgment accuracy when the relationship between the judgments of target traits and targets' self-reported traits could be affective by undue positivity.

Current study. Person judgment research based on restricted presentations of other people (i.e. photographs) has clear advantages to replicability and efficiency. The alternative paradigms, such as using round-robin groups of interaction (i.e. Albright, et al, 1988; Biesanz et al., 2011; Brown & Benrieir, 2017; Paulhus & Reynolds, 1995; Satchell & Pearson, 2017) are time consuming and resource intensive. However *in vivo* interaction studies are a more realistic presentation of encountering another person for the first time. Thus, the current study investigates the relationship and differences between photograph and post-interaction person and personality judgments. Person judgment research frequently focuses on accuracy at judging the Big Five (i.e. Biesanz et al., 2011; Naumann, et al., 2009) and psychopathic personality (i.e. Holtzman, 2011; Satchell & Pearson, 2017) and these traits will be used in this study. It could be the case that there will be consistency in judgments between photograph and in-person judgment. This would provide reassurance to the

efficient and replicable photograph research. Alternatively, person judgments could be changed by the ability to engage in active perception with the unknown other. It is hypothesised that there will be a change in the person and personality judgment ratings made after interacting with a person previously seen as a photograph. It is not clear in which directions the judgments will change (towards more positive or negative perception), but it is expected that personality judgments would become more accurate.

Method

Participants. A total of 101 pairs of participants took part in the study. Participants were recruited separately by research assistants who sampled participants who should not have met (for example, from various courses across the university). No pairs of participants reported knowing each other after taking part in the study. Due to incomplete videos and missing data, 97 pairs (N = 194) were retained for the study. The collection of approximately 100 pairs was chosen to be as close to the required sample size (N =176) for a power of .80 with the average effect size in social psychology (r = .21, Richard et al., 2003). All of the sample were undergraduate students (M_{Age} = 20.24, SD_{Age} = 2.64, Female = 108, Male = 86) and were reasonably split between same sex (46%) and other-sex (54%) pairings. There were no consistent or notable effects of dyads being same sex or other-sex pairings on personality judgments, with the largest difference being a small change in Extraversion judgments after interaction (t(190)= 2.02, p(uncorrected)= .045, d= .30). This study did not ask participants to self-report ethnicity.

Materials.

Personality measures. Participants completed the 44-item Big Five Inventory (John, Naumann & Soto, 2008) and the 58-item Triarchic Psychopathy Measure (TriPM, Patrick, 2010) as assessments of their personality traits. We detail the scales and assess the reliability of the factors using average Intraclass Correlations (ICC) below.

The Big Five Inventory measures Conscientiousness (diligence, tidiness) with nine items (*ICC* = .74, 95% *CI* [.68, .79]), Agreeableness (friendliness, sociability) with nine items (*ICC* = .66, 95% *CI* [.59, .73]), Neuroticism (anxiety) with eight items (*ICC* = .79, 95% *CI* [.75, .83]), Openness

(creativity, aesthetic interests) with 10 items (*ICC* = .71, 95% *CI* [.65, .77]) and Extraversion with eight items (*ICC* = .69, 95% *CI* [.62, .75]). Responses are gathered on a scale of *Disagree Strongly* (1) to *Agree Strongly* (5) on person descriptions such as "I am someone who is talkative".

The TriPM measures three domains of psychopathic personality. There are 19 items to assess Boldness (*ICC* = .80, 95% *CI* [.75, .84]) a measure of confidence, risk-taking and sensation seeking. Disinhibition is assessed with 20 items *(ICC* = .79, 95% *CI* [.75, .83]) detailing impulsivity and poor self-control. Finally, 19 items report Meanness (*ICC* = .76, 95% *CI* [.72, .81]) the antisocial facet of the TriPM. Responses are gathered on a scale of *False* (0) to *True* (1) in response to descriptive statements such as "How other people feel is important to me".

Judgments. There are two typical methods of gathering judgments in personality judgment research; social-named and trait-named ratings. Participants completed the social-named judgments and then the trait-named judgments in this study. They completed both of these ratings after viewing the photograph and after the interaction.

'Social-named' ratings is used as a term to refer to more everyday judgments of traits. Participants first rated their partners on rating scales based on everyday language, designed to imitate a more naturalistic presentation of the personality traits. For example, Carney, Colvin and Hall (2007) asked their participants to judge how "moody, talkative, curious" (p. 1063) they perceived target individuals to be. They rated their partners on a scale of *[not judgment]* (1) to *[judgment]* (9). The social-named judgments were *Organised* (reflecting high Conscientiousness and low Disinhibition), *Friendly* (high Agreeableness and low Meanness), *Anxious* (Neuroticism), *Creative* (Openness), *Energetic* (Extraversion) and *Thrillseeking* (which is a term associated with Boldness; Patrick et al., 2009, p.926). This is not only important for following the extant literature, but also it is the case that asking participants to infer how another person responded to a personality measure (see trait-named judgments description below) is a different task to general social perceptions of another person. Trait-named judgments are a perspective taking exercise, where the judges have to, not only remember the trait definitions, but also consider how another person may think of themselves. Instead, social named judgments are possibly more efficient, simply reflecting on a more organic, natural perception of their

dyad partner. We further asked for social-named ratings which could affect interactions, asking for their judgments of how Threatening, Aggressive, Confident and Likeable they perceived their partner to be.

'Trait-named' judgments are when participants are given brief definitions of the academic usage of the personality traits domains (for example Neuroticism is 'anxiety, worry') and judges are asked to rate how they think the target scored on the measure. For example, Fowler, Lilienfeld, and Patrick (2009) asked participants "the degree to which the target […] matched one- to two sentence descriptions of overall psychopathy, Factor 1 (affective and interpersonal) psychopathy features, and Factor 2 (antisocial and criminal lifestyle) psychopathy features" (p.71). Here we ask participants to report their thoughts on whether their partner *Scored Lowly* (1) to *Scored Highly* (9) on each of the Big Five and TriPM domains when given brief definitions (see supplemental materials). Participants had previously completed the personality measures themselves and were aware of how the trait measures were collected.

Coding. To code for the behaviours displayed by participants during the interaction, a team of 13 coders rated the prevalence (from *Not at all* (1) to *Always* (5)) of 76 different behaviours from video recordings of the dyads' interactions. These were the 64 behaviours detailed in the Riverside Behavioural Q-Sort (Funder et al., 2000) with 12 additional items added by the research team on watching the pairs engaging, and informal comments from participants as they explained their reasoning behind their judgments in debriefing with the research team. These were features such as coding for fashionable clothing, shared body space ('head is oriented towards partner') and self-grooming behaviours. The complete list of items can be found in the supplemental materials. It is to be noted that this is not the intended use of the Q-sort behaviours, but the listed behaviours create a good framework for personality-behaviour research.

All 13 coders rated an overlapping subset of the pairs (16 pairs, 32 participants). The consistency of the coders was high when tested on a same subset of 16 pairs (average *ICC* = .87, 95% CI [.86, .88]). All pairs were rated by at least four coders and we retain the median coding for analysis.

For efficiency, an unrotated principal components analysis was used to define a summary of the 76 criteria. A 15-factor solution provided the last solution with an eigenvalue greater than 1 (eigenvalue = 1.06, 72.35% of variance explained). However, the purpose of a principal components analysis is to provide efficiency of analysis and a 15-factor solution does not provide this. Further, this solution did not provide a clear summary of behaviours (some factors had no best-loadings). On observing the scree plot (see supplemental materials), the curve of eigenvalues normalises at three factors (eigenvalue = 3.59, 44% of variance) with more components only providing minor improvements. We retain regression-derived factor scores for a three factor solution for analysis. The first factor summarised 'Engaging' behaviour, being most strongly associated with coding a participants' enjoyment, interest, likeability and energy in the interaction (see supplemental materials). The second factor was related to insecurities in the interaction, irritability, hostility and seeking advice from interacting partner. We summarised this factor as 'Insecure' behaviour. The third factor can be summarised as 'Dominant' behaviour, relating to participants being more competitive, taking a position of superiority and initiating physical contact in the interaction.

Procedure. Participants arrived at the lab separately and were told to meet the researchers at two different rooms. Both of these rooms connected to a third space where the interaction would take place. On arriving and providing informed consent, a research assistant took a photograph of the participants' faces (from shoulders upwards), holding a neutral expression. The neutral expression was chosen to keep the taken photograph as close to the extent literature precedent as possible, it should be noted that some participants struggled to hold a neutral face. Then each participant began completing the Big Five and TriPM personality measures on paper. Whilst participants were completing the psychometric measures, the research assistants emailed each other the photographs. The photograph of participant 'A' was sent to the research assistant working with participant 'B', and the photograph of participant 'B' was sent to the research assistant working with participant 'A'. After completing the personality measures, the participants were shown the photograph of their to-be partner and made their social and trait judgments (henceforth Time 1). When both participants completed this, they were informed that they would now be engaging in a short interaction. In the

main study room was a rectangle marked on the floor with white tape (approximately 250x100cm, see figure 1). This space was marked so that participants had the most space possible for interaction, whilst still being in range of the cameras for video recording of the dyads. Participants were given limited instructions for how they spent their time in the room. They were told (verbatim):

"The only thing we'd like you to do is to head into the room and stand between the white line and the window. There you will interact with another person. It is up to you what you do in that space but bear in mind that we will be asking your first impressions of that person after you have completed the study. You will be allowed to stay in the room for a maximum of five minutes. We would highlight that you can leave whenever you like, by simply leaving through this door."

It is worth highlighting that these instructions were designed with giving the participants the upmost freedom in how they may choose to interact in the space. One dyad chose not to talk during this time, one dyad chose to move chairs into the interaction space but most dyads used this space to stand and talk. We did not prescribe any behaviour to the participants in this study, but the instructions do remind participants that they will be later making judgments of their interacting partner.

Only 16 pairs left the session early, and the rest of the pairs were asked to return to their individual rooms after five minutes. Wall-mounted cameras in the study room were placed and moved in real-time so that the both participants could be seen face-on for later coding. Participants returned to their own rooms and completed the same ratings (henceforth Time 2). They were then thanked for their time and debriefed.

Figure 1. A photograph of the 250 x 100cm marked area for interaction. It should be noted that one dyad chose to move the chairs into the interaction area for their session.

Analysis. Linear mixed models will be used to estimate the consistency between Time 1 and Time 2, including random factors for the dyads to assess for the amount of variation between pairs when making judgments (Kuznestova, Brockhoff & Christensen, 2017). It should be noted that the random factor variation may be difficult to interpret with 97 levels (dyads) and $n = 2$ per level. Within-subjects' Cohen's *d* will be used to estimate the difference in Time 1 to Time 2 ratings. Personality judgment accuracy will also be tested by correlating the personality traits of a participant with the ratings they received for both Time 1 and Time 2. A comparison in the size of these correlations can be conducted using William's *T* (Steiger, 1980). This tests for the difference between two correlations with a shared dependent variable (target traits). Further, Cross-Intraclass Correlation (CICC) will be used to assess the accuracy of judgments whilst accounting for shared dyad personality, rating behaviour and the judge's own traits on perception (see Rogers, Wood & Furr, 2018). The CICC is analysed for size of effect using a standard *z* output.

An unrotated principal components analysis will be used to investigate how the social judgments shared variance, imitating previous similar studies (i.e. Oostergof & Todorov, 2008; Sutherland et al., 2018; Walker & Vetter, 2009; Wang, et al., 2016; Wang et al., 2018). It is of interest to see if the judgments share variance in the same manner before and after the interaction.

Further correlations will demonstrate the relationship between the behaviours exhibited by the participants with the Time 1 and Time 2 judgments and participant self-reported traits. The relationship between coded behaviours and personality could be due to reciprocal interpersonal activity (i.e. Extraversion brings out Extraversion in an interacting partner), thus CICCs will again be used to attempt for these cross-dyad effects. Cases of missing data will be treated as missing and not replaced with computed variables.

Results. The data for this study are all open access on the Open Science Framework (https://osf.io/bx8zv/).

Judgment consistency and change. It should be noted that there were the expected correlations between the social- and trait-named judgments which were expected to access the same traits (i.e. Creative and Openness) but these were moderate correlations at best (Mean correlation = .33 largest correlation: Agreeableness and Friendliness r = .57).

First, the agreement and divergence for the trait- and social-named judgments is evaluated. Table 1 presents the consistencies and differences between Time 1 (after viewing photograph) and Time 2 (after interaction) in trait-named judgments, and Table 2 presents this information for social-named judgments. In almost all cases, judgements at Time 1 correlated with Time 2; a target who was rated as more [judgment] at Time 1 was also rated as more [judgment] at Time 2. However, the ratings were also different on average. The changes were typically towards the pro-social, such as judgments for Agreeableness and likeable becoming more agreeable and likeable and judgments of Meanness and threatening becoming less mean and threatening (Figure 2).

Trait-named judgments of Neuroticism and social-named judgements of Anxiety and Creativity (and to a certain extent Organisation) did not change after interaction. The random effects

of the dyads were near zero in almost all cases, with the exception of Extraversion and Boldness (in line with the expressivity halo; Bernieri et al., 1996).

Table 1. *The consistency and difference in trait-named judgments made by participants after viewing a photograph (Time 1) and after interacting (Time 2) with their partner*

	Mean judge rating (SD)		Mixed model test for consistency[b]		
Trait	Time 1	Time 2[a]	Consistency (s.e.)	Dyad (*SD*)	Difference (*d*)
Conscientiousness	5.74 (1.57)	6.21 (1.61)	0.37 (0.07)***	0.00 (0.00)	0.27***
Agreeableness	6.38 (1.44)	7.06 (1.39)	0.31 (0.07)***	0.05 (0.22)	0.41***
Neuroticism	4.19 (1.68)	4.25 (2.04)	0.27 (0.06)***	0.11 (0.33)	0.03
Openness	5.61 (1.70)	5.35 (1.71)	0.26 (0.07)***	0.15 (0.39)	0.13***
Extraversion	5.58 (1.70)	6.19 (1.73)	0.34 (0.07)***	0.21 (0.45)	0.31***
Boldness	5.59 (1.68)	6.01 (1.88)	0.33 (0.06)***	0.37 (0.60)	0.21**
Meanness	3.02 (1.69)	2.24 (1.38)	0.32 (0.09)***	0.00 (0.00)	0.41***
Disinhibition	4.25 (1.80)	4.19 (1.96)	0.32 (0.06)***	0.00 (0.00)	0.03

*p< .05, **p<.01, ***p< .001
[a]Due to incomplete data for Time 2, N=190.
[b]Mixed models built using formula fit <- lmer(Time1 ~ Time2 + (1|Dyad)) where time change is a fixed effect and with a random effect of dyad
Significance provided for *d* is computed using a paired-samples t test.

Table 2. *The consistency and difference in social-named judgments made by participants after viewing a photograph (Time 1) and after interacting (Time 2) with their partner*

	Mean judge rating (SD)		Mixed model test for consistency[b]		
Social judgment	Time 1	Time 2[a]	Consistency (s.e.)	Dyad (*SD*)	Difference (*d*)
Organised	5.56 (1.76)	5.88 (1.93)	0.38 (0.06)***	0.00 (0.00)	0.16*
Friendly	6.36 (1.90)	7.54 (1.71)	0.38 (0.08)***	0.00 (0.00)	0.57***
Anxious	4.23 (1.82)	4.45 (2.11)	0.15 (0.06)*	0.00 (0.00)	0.09
Creative	5.62 (1.50)	5.87 (1.65)	0.30 (0.06)***	0.00 (0.00)	0.14
Energetic	5.26 (1.81)	6.06 (1.80)	0.32 (0.07)***	0.36 (0.60)	0.38***
Thrillseeking	4.91 (1.78)	5.39 (1.78)	0.38 (0.07)***	0.00 (0.00)	0.24**
Confident	5.81 (1.66)	6.36 (1.89)	0.33 (0.06)***	0.00 (0.00)	0.28***
Threatening	2.25 (1.46)	1.64 (0.99)	0.21 (0.11)*	0.00 (0.00)	0.37***
Aggressive	2.27 (1.38)	1.70 (1.09)	0.27 (0.09)**	0.00 (0.00)	0.36**
Likeable	6.52 (1.49)	7.47 (1.41)	0.42 (0.07)***	0.00 (0.00)	0.60***

*p< .05, **p<.01, ***p< .001
[a]Due to incomplete data for Time 2, N=190.
[b]Mixed models built using R formula: lmer(Time1 ~ Time2 + (1|Dyad))
Significance provided for *d* is computed using a paired-samples t test.

Personality judgment accuracy. In person judgment research, accuracy can be defined as the relationship between perceptual judgments and the partner's self-reported personality traits (Funder, 2012). In this study, there was no consistent evidence of personality judgment accuracy at Time 1 in this study using trait-named judgments (Table 3). There was a notable correlation between Time 1 judgments of Creativity and target Openness (see Table 4, Figure 2) and slightly weaker evidence that judgments of Thrillseeking detected target's Boldness, but these were exceptions.

Table 3. *The accuracy of trait-named judgments at detecting target's personality traits after viewing a photograph (Time 1) and after interacting (Time 2) with their partner*

	Raw accuracy (*r*)		Difference	CICC Accuracy (*z*)	
Trait	Time 1	Time 2[a]	William's *T*	Time 1	Time 2[a]
Conscientiousness	.00	.05	0.62	0.02	0.84
Agreeableness	-.11	.07	2.11*	-1.65*	0.93
Neuroticism	.01	.27***	3.17**	1.28	3.85***
Openness	.07	.17*	1.14	0.99	2.27*
Extraversion	.11	.22**	1.34	1.50	2.76**
Boldness	.09	.19**	1.22	1.20	2.34*
Meanness	.07	.07	0.00	0.92	0.91
Disinhibition	.03	.08	0.60	0.40	1.06

*p< .05, **p<.01, ***p< .001
[a]Due to incomplete data for Time 2, N=190.
Supplemental Table A provides mixed models controlling for same sex or not dyads with these effects. The pattern of results is highly similar.
CICC = Cross-Intraclass Correlation given by Rogers et al.'s (2018) code

At Time 2, personality judgment improved (and notably so in some William's *T* tests). Judgments of Neuroticism and Anxiety indicated partner's Neuroticism, Extraversion judgments indicated partner's Extraversion and judgments of Creativity indicated partner's Openness. Boldness was accurately rated in judgments of Boldness and Thrillseeking. Many other correlations improved in size (according to William's *T*) but were still not of noteworthy accuracy (such as for Meanness and Agreeableness). In general interaction improved personality judgment but accuracy was still generally poor.

The CICCs and the raw correlation results were highly similar. Figure 2 summarises judgment accuracy at Time 1 and Time 2.

Table 4. *The accuracy of social-named judgments at detecting target's personality traits after viewing a photograph (Time 1) and after interacting (Time 2) with their partner*

	Raw accuracy (*r*)	Difference	CICC Accuracy (*z*)

Judgment	Target trait	Time 1	Time 2[a]	William's *T*	Time 1	Time 2[a]
Organised	Conscientiousness	-.04	.11	1.92	-0.74	1.75*
Friendly	Agreeableness	-.05	.06	1.33	-0.77	0.91
Anxious	Neuroticism	.13	.19*	0.65	2.03*	2.65**
Creative	Openness	.28***	.20*	0.99	4.06***	2.96**
Energetic	Extraversion	.12	.21**	1.07	1.58	2.66**
Thrillseeking	Boldness	.18*	.20**	0.25	2.73**	2.66**
Friendly	Meanness	-.02	-.11	3.59***	-0.61	-1.45
Organised	Disinhibition	.03	-.08	3.93***	0.82	-1.22

*p< .05, **p<.01, ***p< .001

[a]Due to incomplete data for Time 2, N=190.

Supplemental Table B provides mixed models controlling for same sex or not dyads with these effects. The pattern of results is highly similar.

CICC = Cross-Intraclass Correlation given by Rogers et al.'s (2018) code

Principal components of judgments. Other research has summarised the variance in multiple person judgments as order factors such as perceived 'valence' and 'dominance' (Oostergof & Todorov, 2008) or 'approachability' and 'capability' (Wang et al., 2018) of other people. Following the method used in these previous studies, an unrotated principal components analysis was conducted on the social judgments at Time 1, and then a second analysis was conducted at Time 2 (see Table 5). In this study, using the same eigenvalue selection criteria to suggest number of factors, a three factor solution was most appropriate at Time 1 and at Time 2. The three factors emerged as 'Positive' ratings (i.e. Organised, Friendly, Energetic), 'Negative' ratings (i.e. Threatening, Aggressive) and Anxiety (containing Anxious alone). The component loadings were largely consistent between Time 1 and Time 2, with the notable exception of Confidence and Thrillseeking. These were initially loading with Negative factors when based on judgments made of the photographs, however after interaction these were loading with the Positive domain. This shift is important to note as this relates to the 'approachability' and 'valence' attributes referred to in previous research.

Table 5. *Principal components analysis results showing the loadings of the judgments at Time 1 and Time 2.*

	Time 1			Time 2[a]		
	'Positive'	'Negative'	'Anxiety'	'Positive'	'Negative'	'Anxiety'
Threatening	-.56	**.67**	.26	-.28	**.87**	.12
Aggressive	-.55	**.72**	.28	-.19	**.88**	.17
Organised	**.56**	-.09	.40	**.45**	-.14	.37
Friendly	**.81**	-.18	.04	**.74**	-.26	.20
Anxious	-.01	-.17	**.86**	-.43	-.12	**.78**
Creative	**.64**	.28	.21	**.62**	.12	.33
Energetic	**.69**	.38	.03	**.80**	.30	-.05

Confident	.49	**.69**	-.22	**.82**	.26	-.22
Thrillseeking	.35	**.64**	-.15	**.60**	.41	.01
Likeable	**.82**	-.05	.07	**.80**	-.19	.13

Notes
Top loading factor in each analysis is highlighted in bold.
Time 1 produces a three factor solution with an eigenvalue= 1.17, explaining 68.15% of variance.
Time 2 produces a three factor solution with an eigenvalue= 1.01, explaining 67.75% of variance
[a]Due to incomplete data for Time 2, *N*=190.

Behaviour, judgments, and traits. Participants' behaviours in the interaction were coded on a series of items, summarised as Energetic, Insecure, and Dominant behaviours. Primarily the coding allows an investigation into what behaviours affect Time 2 judgments. The relationship between Time 1 judgments and coded behaviours were analysed, demonstrating any ability of the photograph judgments to predict behaviours. William's *T* comparisons between the Time 1 and Time 2 correlations demonstrate the extent to which behaviour changed perceptions.

In general, Time 1 judgments did not relate to coded behaviour in the study, suggesting that trait- (Table 6) and social- (Table 7) named judgments from photographs did not *anticipate* the behaviour of the partner. At Time 2 many judgments were related to the Engaging behaviours demonstrated by partners, with no general evidence that Insecure or Dominant behaviours correlated with judgments (see Tables 6 and 7). Overall, these Engagement behaviours, such as talkativeness, energy, smiling, laughing, playfulness, questioning and fluency (see supplemental material), were related to generally more prosocial judgments at Time 2. Engaging behaviours increased ratings of Agreeableness, Friendliness, Openness, Extraversion, Energetics, Boldness, Thrillseeking and Likeability (but also Disinhibition) and ratings of Neuroticism and Meanness decreased a little.

Table 6. *Correlations between received trait-named judgments made after viewing a photograph (T1) and after interacting (T2) with their partner and the coded behaviours in the interaction with William's T (T) comparisons of the difference*

	Engaging Behaviour			Insecure Behaviour			Dominant Behaviour		
Trait	T1	T2[a]	*T*	T1	T2[a]	*T*	T1	T2[a]	*T*
Conscientiousness	.08	.09	0.12	-.06	-.02	0.49	.04	-.10	1.74
Agreeableness	.05	.31***	3.17***	-.03	-.17*	1.65	.03	-.03	0.70
Neuroticism	.03	-.15*	2.15*	.07	.13	0.71	-.05	.09	1.66
Openness	.06	.31***	2.96***	-.09	-.08	0.11	.07	-.04	1.25
Extraversion	.03	.27***	2.98***	-.05	.00	0.60	.05	-.06	1.32
Boldness	-.02	.28***	3.80***	.01	-.11	1.46	.07	-.12	2.32*

Meanness	-.12	-.18*	0.69	.15*	.03	1.37	-.04	.04	0.91
Disinhibition	.03	.27***	2.98***	.07	.07	0.00	.01	.09	0.96

*p< .05, **p<.01, ***p< .001
[a]Due to incomplete data for Time 2, N=190.

1

Table 7. *Correlations between received social-named judgments made after viewing a photograph (T1) and after interacting (T2) with their partner and the coded behaviours in the interaction with William's T (T) comparisons of the difference*

	Engaging Behaviour			Insecure Behaviour			Dominant Behaviour		
Social Judgment	T1	T2[a]	*T*	T1	T2[a]	*T*	T1	T2[a]	*T*
Organised	-.00	.16*	2.06*	-.01	-.06	0.64	-.09	-.19*	1.29
Friendly	.06	.27***	2.61*	-.05	-.01	-0.05	.00	-.08	0.96
Anxious	.08	-.07	1.60	.11	.08	0.32	-.08	.08	1.71
Creative	.04	.18	1.68	.02	.01	0.12	.01	.00	0.12
Energetic	.02	.30***	3.42***	-.04	-.10	0.70	.05	.00	0.58
Thrillseeking	.04	.25**	2.66*	-.02	-.00	0.25	.08	-.02	1.23
Confident	.10	.29***	2.43*	-.08	-.11	0.37	.11	-.08	2.37*
Threatening	-.07	-.13	0.63	.11	-.04	1.58	.11	.15*	0.42
Aggressive	-.05	-.06	0.11	.13	.01	1.32	.10	.15*	0.55
Likeable	.08	.36***	3.74*	.02	-.05	0.88	-.04	-.03	0.12

*p< .05, **p<.01, ***p< .001

2

3 Table 8 demonstrates the relationship between self-reported traits and behaviours. We find
4 small, weak evidence that Extraversion, Agreeableness and Boldness related to more Engaging
5 behaviour. Thus it could be the case that Engaging behaviours facilitated Extraversion personality
6 judgment accuracy, but it is more likely that the behaviours supporting person judgment accuracy
7 were not captured in the coding in this study.

8 Figure 2 summarises the relationships between behaviour, judgments at Time 1, judgments at
9 Time 2 and the coded behaviours.

10

Table 8. *Correlations between participant self-reported trait and the coded behaviour during the interaction*

	Engaging Behaviour		Insecure Behaviour		Dominant Behaviour	
Trait	Raw *r*	CICC (*z*)	Raw *r*	CICC (*z*)	Raw *r*	CICC (*z*)
Conscientiousness	.07	0.76	-.13	-1.41	-.07	-0.72
Agreeableness	.15*	1.56	.01	0.08	.04	0.45
Neuroticism	-.10	-1.04	.04	0.36	-.02	-0.25
Openness	.07	0.76	-.01	-0.10	.01	0.05
Extraversion	.20**	2.03*	-.06	-0.65	-.10	-0.10
Boldness	.15*	1.61	-.11	-1.16	.01	0.10
Meanness	-.11	-1.11	-.05	-0.52	-.02	-0.16
Disinhibition	-.08	-0.79	-.05	-0.55	.04	0.41

*p< .05, **p<.01, ***p< .001

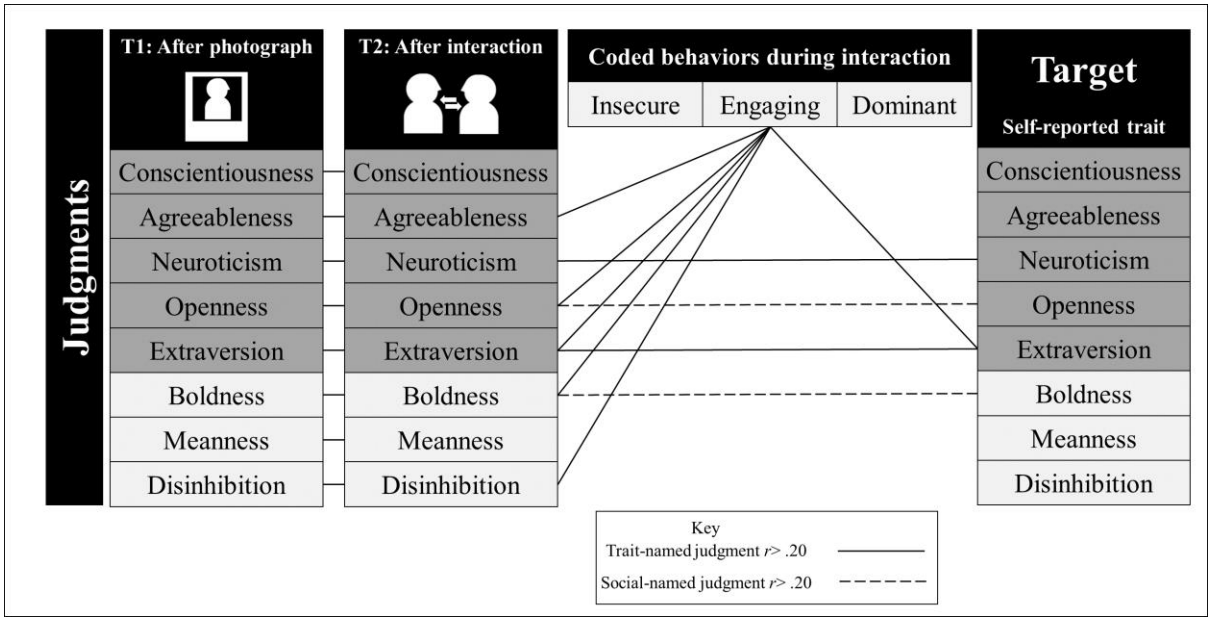


Figure 1. A visual summary of the results from tables 1, 2, 4 and 5. Only correlations larger than .20 between traits, behaviours Time 1 (T1) judgments and Time 2 (judgments) are presented for efficiency. Readers should consult the tables for exact sizes of effects.

Discussion

This study investigated the effect of interaction on person and personality judgment. Participants first rated interacting partners from observing a photograph of their face and then rated their partners again after a brief (less than five minute) interaction. Judgments before and after interaction were consistent but different. A target rated more friendly at Time 1 (before interaction) was still considered more friendly at Time 2 (after interaction), but was more strongly rated as such. Although there was notable consistency between Time 1 and Time 2 judgments, the stability of the ratings was only moderate (average stability correlation r = .31) and therefore only approximately 10% of the variance in Time 2 ratings can be explained with Time 1 ratings (on average). It was possible to account for some of the variance in changed ratings by coding for behaviours in the interaction. It was primarily Engaging behaviours such as smiling, acting interested, being energetic, being relaxed and being talkative that related to more positive person judgments.

These results can be interpreted in two ways. First, we have some evidence that judgments of other people from photographs show some stability for impressions after interaction – photographs of faces are an efficient proxy for studying interpersonal interaction. However, this stability is arguably small, and the pre-interaction judgments were more negative than the more pro-social judgments made after interaction. This is similar to the increased liking seen in other 'get-acquainted interaction' studies (see Finkel et al., 2015).

Secondly, the current results could be seen to support the evidence that interaction changes perceptions. Personality judgments were more accurate at Time 2, when participants had the opportunities to be active perceivers (Good, 2007). Everyday person judgment involves conversational probes and (re)action with a partner. This relates to the Uncertainty Reduction Theory that posits that increases in verbal and non-verbal communication decreases uncertainty about our perceptions of others (Sunnafrank, 1986). The more pro-social ratings observed at Time 2, could be the sign of participants using the interaction to gain more certainty about their partners.

It should be noted that there is an unavoidable order-effect in this study. Participants were exposed to the photograph of the partner before interaction. This is an inevitable consequence of the study design as participants could not be provided with a more information-rich (*in vivo*) exposure followed by an information-poor (photograph) exposure. As such, it could be the case that seeing the photograph of the partner may shape the following interaction. It is possible that without the photograph 'priming' expectations of the interaction, judgments could be further different. Future research could randomly either see a photograph of their to-be partner or not before an interaction, to investigate if this methodological confound affects studies of this type.

In person judgment studies it is most ideal to account for variance in 'good' judges and 'good' targets (Funder, 1999). Some judges of personality will be better at *detecting* and *utilizing* (in RAM terms) traits than others (see Funder, 2012; Powell & Bourdage, 2016; Schmid Mast, Bangerter, Bulliard & Aerni, 2011; Satchell, Morris, Akehurst & Morrison, 2017). Some targets will make their *relevant* personality behaviours more *available* to be detected (see Funder, 1999). This is what Benrieri et al. (1996) refer to as the 'expressivity halo', and the present results provide further support

to the ease of detecting extraversion over other traits (Albright et al., 1988; Bernieri et al., 1994; Bernieri et al., 1996; Funder & Colvin, 1988). This was the case in the current study where it was Extraversion and Boldness judgments that were most influenced by variation in dyads, suggesting that there is an element of reciprocity affecting perceptions of expressivity. Research could further explore judge and target variance using round-robin paradigms (Albright et al., 1998; Cronbach, 1955; Kenny & Albright, 1987; Warner, Kenny & Stoto, 1979). Brown and Bernieri (2017) do this well in their study of developing first impressions, by following interacting polyads of six to eight individuals over a series of weeks. By doing this they were best able to account for judges' response patterns and targets' salience of traits. In the present investigation accounting for this variance was not possible, and so the results may be affected by the judges' rating preferences, the judges' ability to detect traits and the salience of the targets' traits. Future research could better distil person judgment accuracy of photographs and later interaction using a round-robin method where multiple actors would rate multiple targets. This method would synthesise the current work with the work of other person-judgment researchers (such as Brown & Bernieri, 2017). A study of this type would be logistically difficult, but would offer key theoretical insight and statistical control.

This paper opened with a quote from Neisser's (1980) critique on the use of "passive onlookers" (p.603) in social perception research. The current study was an effort to adjust this particular issue. However, there is a second point raised in Neisser's piece. He comments on the usage of overt rating scales, evaluating a target person on a (perhaps) 1-9 scale:

"*Who actually behaves in this way? Not everybody: in some societies and social groups, hardly anybody. Of course, it might – and has been argued- that everyone makes judgments covertly […] But I know of no evidence to support such a view, and it strikes me as prime facie implausible*" (p.604).

It is highly unlikely that individuals make such concrete ratings of others in day-to-day life. Such judgments may manifest as feelings or likings towards others and, as such, the current activity of putting to paper how "*Good* (9) to *Bad* (1)" an other person is unusual. As such, the current research

falls short of another ecological psychology target; *action fidelity* (Stroffregen, Bardy, Smart & Pagulayan, 2003). The activities participants engage in should be reflective of the activity we wish to study. Arguably, paradigmatic change is needed to find a way of studying first impressions in a naturalistic, yet well powered and replicable manner. In the current work, the ecological qualities of active perception (Good, 2007) and representative design (Brunswick, 1956) have been addressed, but further work should consider improving the action fidelity of *in vivo* person perception.

Summary. The current study suggests that social- and trait-named judgments of first impressions from photographs show some consistency with judgments made after first interactions. The accuracy of personality judgments were improved after a short interaction, with opportunities to engage in active perception, but even these judgments were still generally poor. This study highlights the importance of the ecological validity (representative design) of the methodology and how proxies for first interaction changes conceptual understanding of first impressions. Whilst photographs have their place in first impressions, it is still the case that most first contact with another person involves 'mixing it up with folks' (to paraphrase Neisser) and research should reflect this. More work needs to be done to understand the issue of 'bubble-ism' in person perception research where the studies focus on one body feature at a time, such as face width, eye colour, hair style, etc, rather than part of a complete whole. Much more methodological and theoretical work is needed to increasingly naturalise research on 'first impressions'.

References

Adamson, L. B., & Frick, J. E. (2003). The Still Face: A History of a Shared Experimental Paradigm. *Infancy*, 4(4), 451–473. doi:10.1207/s15327078in0404_01

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.

Araújo, D., Davids, K., & Passos, P. (2007). Ecological Validity, Representative Design, and Correspondence Between Experimental Task Constraints and Behavioral Setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology*, *19*(1), 69–78. doi:10.1080/10407410709336951

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, 78, 34–42. doi:10.1016/j.jesp.2018.04.011

Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology*, *71*(1), 110–129. doi:10.1037/0022-3514.71.1.110

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. doi:10.1177/0146167294204004

Biesanz, J. C., Human, L. J., Paquin, A.-C., Chan, M., Parisotto, K. L., Sarracino, J., & Gillis, R. L. (2011). Do We Know When Our Impressions of Others Are Valid? Evidence for Realistic Accuracy Awareness in First Impressions of Personality. *Social Psychological and Personality Science*, *2*(5), 452–459. doi:10.1177/1948550610397211

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*(4), 703-706.

Brown, J. A., & Bernieri, F. (2017). Trait Perception Accuracy and Acquaintance Within Groups: Tracking Accuracy Development. *Personality and Social Psychology Bulletin*, 43(5), 716–728. doi:10.1177/0146167217695557

van Breen, J. A., De Dreu, C. K., & Kret, M. E. (2018). Pupil to pupil: The effect of a partner's pupil size on (dis) honest behavior. *Journal of Experimental Social Psychology*, *74*, 231-245. doi: 10.1016/j.jesp.2017.09.009

Bruce, V., & Young, A. W. (2012). *Face perception*. London, England: Psychology Press

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

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*(5), 1054–1072. doi:10.1016/j.jrp.2007.01.004

Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behavior in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society of London, Series B:Biological Sciences, 275*, 2651–2656. doi:10.1098/rspb.2008.0873

Carré, J. M., Morrissey, M. D., Mondloch, C. J., & McCormick, C. M. (2010). Estimating Aggression from Emotionally Neutral Faces: *Which Facial Cues are Diagnostic? Perception, 39*(3), 356–377. doi:10.1068/p6543

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70*(2), 205–218. doi:10.1037/0022-3514.70.2.205

Costa, M., Lio, G., Gomez, A., & Sirigu, A. (2017). How components of facial width to height ratio differently contribute to the perception of social traits. *PLOS ONE*, 12(2), e0172739. doi:10.1371/journal.pone.0172739

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52*(3), 177–193. doi:10.1037/h0044919

Deska, J. C., & Hugenberg, K. (2018). Targets' facial width-to-height ratio biases pain judgments. *Journal of Experimental Social Psychology, 74*, 56-64. doi:10.31234/osf.io/xcbnh

Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology, 114(1),* 75–94. doi:10.1037/pspi0000110

Dixson, B. J., & Vasey, P. L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness. *Behavioral Ecology*, *23*(3), 481-490. doi:10.1093/beheco/arr214

Finkel, E. J., Norton, M. I., Reis, H. T., Ariely, D., Caprariello, P. A., Eastwick, P. W., ... & Maniaci, M. R. (2015). When does familiarity promote versus undermine interpersonal attraction? A proposed integrative model from erstwhile adversaries. *Perspectives on Psychological Science*, *10*(1), 3-19.

Fowler, K. A., Lilienfeld, S. O., & Patrick, C. J. (2009). Detecting psychopathy from thin slices of behavior. *Psychological Assessment, 21*(1), 68–78. doi:10.1037/a0014938

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 1, 177–182. doi:10.1177

Funder, D. C. (Ed.). (1999). *Personality judgment: A realistic approach to person perception*. London: Academic Press.

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), 149–158. doi:10.1037/0022-3514.55.1.149

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside Behavioral Q-sort: A Tool for the Description of Social Behavior. *Journal of Personality, 68*(3), 451-489.

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, 23(5), 369–401. doi:10.1002/per.724

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from Meta-Analyses of the Facial Width-to-Height Ratio as an Evolved Cue of Threat. *PLOS ONE*, 10(7), e0132726. doi:10.1371/journal.pone.0132726

Gonzalez, A. M., Steele, J. R., & Baron, A. S. (2017). Reducing children's implicit racial bias through exposure to positive out-group exemplars. *Child Development, 88*(1), 123-130. doi:10.1111/cdev.12582

Good, J. M. M. (2007). The Affordances for Social Psychology of the Ecological Approach to Social Knowing. *Theory & Psychology*, 17(2), 265–295. doi:10.1177/0959354307075046

Gordon, D. S., & Platek, S. M. (2009). Trustworthy? The brain knows: Implicit neural responses to faces that vary in dark triad personality characteristics and trustworthiness. *Journal of Social, Evolutionary, and Cultural Psychology*, *3*(3), 182-200. doi: 10.1037/h0099323

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychology, Behavior, and Social Networking*, 14(9), 483–488. doi:10.1089/cyber.2010.0087

Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's Facial Width-to-Height Ratio Predicts Aggression: A Meta-Analysis. *PLOS ONE, 10(4),* e0122637. doi:10.1371/journal.pone.0122637

Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, *49*(4), 747-754. doi:10.1016/j.jesp.2013.02.015

Hillman, J., Vrij, A., & Mann, S. (2011). Um … they were wearing …: The effect of deception on specific hand gestures. *Legal and Criminological Psychology, 17*(2), 336–345. doi:10.1111/j.2044-8333.2011.02014.x

Holtzman, N. S. (2011). Facing a psychopath: Detecting the dark triad from emotionally-neutral faces, using prototypes from the Personality Faceaurus. *Journal of Research in Personality, 45*(6), 648–654. doi:10.1016/j.jrp.2011.09.002

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A.

Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.

Jones, A. L., & Kramer, R. S. S. (2016). Facial Cosmetics and Attractiveness: Comparing the Effect Sizes of Professionally-Applied Cosmetics and Identity. *PLOS ONE, 11*(10), e0164218. doi:10.1371/journal.pone.0164218

Jones, A. L., Russell, R., & Ward, R. (2015). Cosmetics Alter Biologically-Based Factors of Beauty: Evidence from Facial Contrast. *Evolutionary Psychology*, *13*(1), 210-229 doi:10.1177/147470491501300113

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*(5), 741–741. doi:10.1037/a0013550

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, *102*(3), 390–402. doi:10.1037/0033-2909.102.3.390

Kim, M-G., & Lee, M-H (2011). A Study on Image Perception according to Wedding Headdress, Hairstyle, and Neckline. *The Research Journal of the Costume Culture*, *19*(5), 981–992. doi:10.29049/rjcc.2011.19.5.981

Kuznetsova A., Brockhoff P.B., & Christensen, R.H.B. (2017).lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13.

Little, A. C. (2014). Facial appearance and leader choice in different contexts: Evidence for task contingent selection based on implicit and learned face-behaviour/face-ability associations. *The Leadership Quarterly*, *25*(5), 865–874. doi:10.1016/j.leaqua.2014.04.002

Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology*, *81*(1), 11–21. doi:10.1037/0021-9010.81.1.11

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, *94*(3), 365–381. doi:10.1037/0022-3514.94.3.365

Morrison, D., Wang, H., Hahn, A. C., Jones, B. C., & DeBruine, L. M. (2017). Predicting the reward value of faces and bodies from social perception. *PLOS ONE, 12*(9), e0185093. doi:10.1371/journal.pone.0185093

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality Judgments Based on Physical Appearance. *Personality and Social Psychology Bulletin, 35*(12), 1661–1671. doi:10.1177/0146167209346309

Neisser, U. (1980). On "Social Knowing." *Personality and Social Psychology Bulletin*, *6*(4), 601–605. doi:10.1177/014616728064012

Norton, M. I., Frost, J. H., & Ariely, D. (2007). Less is more: The lure of ambiguity, or why familiarity breeds contempt. *Journal of Personality and Social Psychology, 92,* 97-105. doi:10.1037/e683162011-051

Norton, M. I., Frost, J. H., & Ariely, D. (2011). Does familiarity breed contempt or liking? Commentary on Reis, Maniaci, Caprariello, Eastwick, and Finkel (2011). *Journal of Personality and Social Psychology, 101,* 571-574. doi:10.1037/a0023202

Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.01516

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 11087-11092. doi:10.1073/pnas.0805664105

Open Science Collaboration. (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science*, 7, 657–660. doi:10.1177/1745691612462588

Ormiston, M. E., Wong, E. M., & Haselhuhn, M. P. (2017). Facial-width-to-height ratio predicts perceptions of integrity in males. *Personality and Individual Differences, 105,* 40–42. doi:10.1016/j.paid.2016.09.017

Patrick, C. J. (2010). Operationalizing the triarchic conceptualizationof psychopathy: Preliminary description of brief scales for assessment of boldness, meanness, and disinhibition. Unpublished test manual, Florida State University, Tallahassee, Florida. Test retrieved from https://www.phenxtoolkit.org/index.php?pageLink=browse.protocoldetails&id=121601

Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology*, *21*, 913. doi:10.1017/s0954579409000492

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

Powell, D. (2016). The Detection of Personality Traits in Employment Interviews: Can "Good Judges" Be Trained? *Personality and Individual Differences*, *101*, 506–507. doi:10.1016/j.paid.2016.05.263

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011a). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology*, *101*, 557-570. doi:10.1037/a0022885

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011b). In live interaction, does familiarity promote attraction or contempt?: A reply to Norton, Frost, and Ariely (2011). *Journal of Personality and Social Psychology*, *101*, 575-578. doi:10.1037/a0023471

Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One Hundred Years of Social Psychology Quantitatively Described. *Review of General Psychology*, *7*(4), 331–363. doi:10.1037/1089-2680.7.4.331

Rogers, K. H., Wood, D., & Furr, R. M. (2018). Assessment of similarity and self-other agreement in dyadic relationships. *Journal of Social and Personal Relationships*, 35(1), 112–134. doi:10.1177/0265407517712615

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology, 44*(4), 1100–1105. doi:10.1016/j.jesp.2007.12.001

Rule, N. O., Bjornsdottir, R. T., Tskhay, K. O., & Ambady, N. (2016). Subtle perceptions of male sexual orientation influence occupational opportunities. *Journal of Applied Psychology, 101*(12), 1687–1704. doi:10.1037/apl0000148

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., … Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage, 54*(3), 2524–2533. doi:10.1016/j.neuroimage.2010.10.011

Satchell, L., & Pearson, D. (2017). The Social Salience of Students' Sub-Clinical Psychopathic Personality. *Current Psychology*. doi:10.1007/s12144-017-9758-6

Satchell, L., Morris, P., Akehurst, L., & Morrison, E. (2017). Can Judgments of Threat Reflect an Approaching Person's Trait Aggression? *Current Psychology*. doi:10.1007/s12144-016-9557-5

Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews?. *International Journal of Selection and Assessment, 19*(2), 198-208. doi:10.1111/j.1468-2389.2011.00547.x

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245–251. doi:10.1037/0033-2909.87.2.245

Stoffregen, T. A., Bardy, B. G., Smart, L. J., & Pagulayan, R. J. (2003). On the nature and evaluation of fidelity in virtual environments. In L. J. Hettinger & M. W. Haas (Eds.), *Virtual and adaptive environments: Applications, implications, and human performance issues* (pp. 111–128). Mahwah, NJ: Lawrence Erlbaum Associates, Inc

Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial First Impressions Across Culture: Data-Driven Modeling of Chinese and British Perceivers' Unconstrained Facial Impressions. *Personality and Social Psychology Bulletin, 44*(4), 521–537. doi:10.1177/0146167217744194

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2014). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology, 106*(2), 186–208. doi:10.1111/bjop.12085

Sunnafrank, M. (1986). Predicted outcome value during initial interactions A reformulation of Uncertainty Reduction Theory. *Human Communication Research, 13(1),* 3–33. doi:10.1111/j.1468-2958.1986.tb00092.x

Talamas, S. N., Mavor, K. I., & Perrett, D. I. (2016). Blinded by Beauty: Attractiveness Bias and Accurate Perceptions of Academic Performance. *PLOS ONE, 11(*2), e0148284. doi:10.1371/journal.pone.0148284

Vazire, S., & Gosling, S. D. (2004). e-Perceptions: personality impressions based on personal websites. *Journal of personality and social psychology*, *87*(1), 123–132. doi:10.1037/0022-3514.87.1.123

Walker, M., & Vetter, T. (2009). Portraits made to measure: Manipulating social judgments about individuals with a statistical face model. *Journal of Vision*, *9*(11), 1–13. doi:10.1167/9.11.12

Wang, H., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2016). The Motivational Salience of Faces Is Related to Both Their Valence and Dominance. *PLOS ONE, 11*(8), e0161114. doi:10.1371/journal.pone.0161114

Wang, H., Han, C., Hahn, A., Fasolt, V., Morrison, D.,... Jones, B. C. (2018, February 23). A data-driven study of Chinese participants' social judgments of Chinese faces. Retrieved from psyarxiv.com/38taf

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, *37*, 1742–1757. doi:10.1037/0022-3514.37.10.1742

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science*, *17*(7), 592-598. doi:10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology, 113*(1), 59–80. doi:10.1037/pspi0000092

Supplemental materials.

Trait definitions provided to participants.

The person in the photo has completed the same personality measures as you did. The personality measures measure the following domains. Please read the definitions with care.

Trait	**Definition**	**Trait**	**Definition**
Conscientiousness	Tidiness, organisation	Boldness	Confident, self-assured
Agreeableness	Sociability, friendliness	Meanness	Manipulator, cheater
Neuroticism	Anxiety, worry	Disinhibition	Spontaneous, uncontrolled
Openness	Creativity, 'art-yness'		
Extraversion	Energetic, outgoing		

Behaviour coding scheme definitions and factor loadings.

Notes.

Items 1 to 64 are those detailed in the Riverside Behavioural Q-Sort (Funder, Furr & Colvin, 2000)

Items 65 to 76 are additional items added by the current researchers.

The factor loadings for the three-factor dimension reduction solution are found in the right hand columns. The process of dimension reduction is found in text.

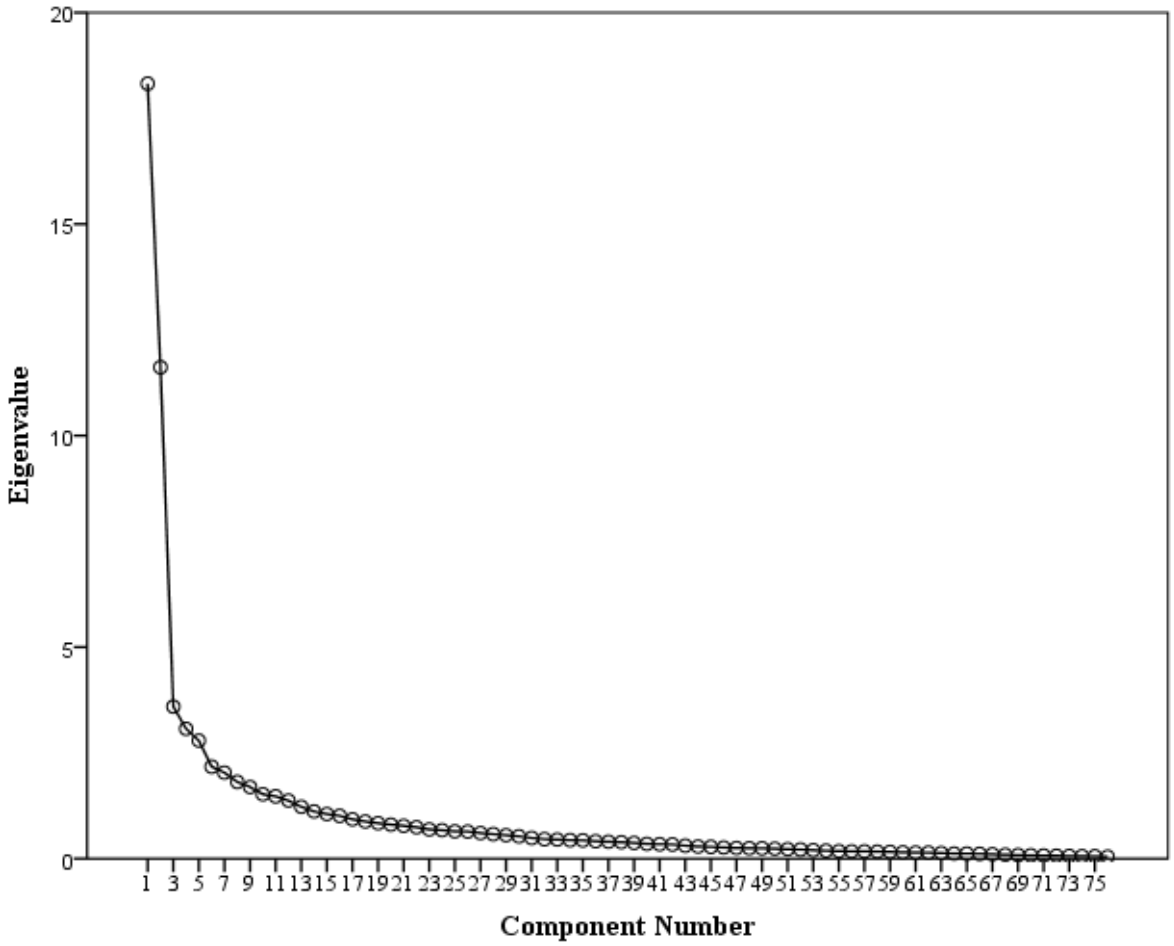
			PCA loading		
No.	Code	Detail	Engaging	Insecure	Dominating
1	Aware	Expresses awareness of being on camera or in an experiment (regardless of whether reaction is positive or negative).	-0.04	0.50	0.03
2	Questioning	Interviews his or her partner (e.g., asks a series of questions).	0.65	-0.09	-0.18
3	Offering	Volunteers a large amount of information about self.	0.77	-0.14	-0.12
4	Interested	Seems interested in what partner has to say.	0.86	-0.24	-0.08
5	Controlling	Tries to control the interaction (disregard whether attempts at control succeed or not).	0.54	0.24	0.02
6	Dominating	Dominates the interaction (disregard intention, e.g., if subject dominates Interaction "by default" because the partner does very little, this item should receive high placement)	0.63	0.17	0.01
7	Relaxed	Appears to be relaxed and comfortable.	0.77	-0.28	-0.11
8	Social	Exhibits social skills (e.g., does things to make partner comfortable, keeps conversation moving, entertains or charms the partner)	0.72	-0.35	-0.20
9	Reserved	Is reserved and inexpressive (e.g., expresses little affect (emotion); acts in a still, formal manner).	-0.34	0.43	-0.25
10	Laughs	Laughs frequently (disregard whether or not laughter appears to be "nervous" or genuine).	0.74	-0.06	-0.01
11	Smiles	Smiles frequently.	0.75	-0.24	-0.07
12	Animated	Is physically animated; moves around a great deal.	0.57	0.23	0.28
13	Liking	Seems to like partner (e.g. would probably like to be friends with partner).	0.71	-0.20	-0.17
14	Awkward	Exhibits an awkward interpersonal style (e.g., seems to have difficulty knowing what to say; mumbles; fails to respond to partner's conversational advances)	-0.54	0.48	-0.15
15	Comparative	Compares self to other(s) (whether others are present or not).	0.47	0.58	-0.18
16	Energetic	Shows high enthusiasm and a high energy level.	0.80	-0.21	0.13
17	Broad	Shows a wide range of interests. (e.g., talks about many topics).	0.67	-0.22	-0.12
18	Monologue	Talks at, rather than with, partner (e.g., conducts a monologue, ignores what partner says).	0.37	0.63	-0.23

19	Agreement	Expresses agreement frequently (high score implies agreement is expressed unusually often, e.g., in response to each and every statement partner(s) makes. Low score implies unusual lack of expression of agreement).	0.61	-0.15	0.19
20	Criticism	Expresses criticism (of anybody or anything; low score implies expresses praise).	0.30	0.39	0.48
21	Talkative	Is talkative (as observed In this situation).	0.76	-0.26	-0.13
22	Insecure	Expresses insecurity (e.g.,seems touchy or overly sensitive}.	-0.12	0.75	-0.17
23	Anxious	Shows physical signs of tension or anxiety (e.g., fidgets nervously, voice wavers). (Lack of signs of anxiety = middle score; low score = lack of signs under circumstances where you would expect to see them.)	-0.35	0.54	0.05
24	Intelligence	Exhibits a high degree of intelligence (NB: At issue is what is displayed in the interaction, not what may or may not be latent. Thus give this item a high score only if subject actually says or does something of high intelligence. A low score implies exhibition of low intelligence; medium score = no information one way or another).	0.46	0.22	0.44
25	Sympathy	Expresses sympathy toward partner (low placement implies unusual lack of sympathy).	0.46	0.21	0.33
26	Humorous	Initiates humour.	0.75	-0.06	0.05
27	Fishing	Seeks reassurance from partner (e.g., asks for agreement, fishes for praise).	0.50	0.64	-0.20
28	Superiority	Exhibits condescending behaviour (acts as if self is superior to partner in one or more ways. Low score implies acting inferior)	0.28	0.33	0.54
29	Likable	Seems likable (to others present).	0.80	-0.29	-0.02
30	Seeks	Seeks advice from partner.	0.29	0.68	-0.13
31	Narcissism	Appears to regard self as physically attractive (nonverbal cues probably will be used to judge this item; examples might include preening, posing, etc.).	0.18	0.59	-0.12
32	Irritable	Acts irritated.	0.00	0.71	-0.10
33	Warmth	Expresses warmth (to anyone, e.g., include any references to “my close friend”, etc.)	0.60	0.33	-0.10
34	Saboteur	Tries to undermine, sabotage, or obstruct (either the experiment or partner)	-0.06	0.60	-0.28
35	Hostile	Expresses hostility (no matter toward whom or what – including the experiment or partner)	-0.15	0.70	-0.19
36	Unusual	Is unusual or unconventional in appearance.	-0.07	0.73	-0.11
37	Timid	Behaves in a fearful or timid manner	-0.28	0.66	-0.21
38	Expressive	Is expressive in face, voice, or gestures.	0.70	-0.05	0.06
39	Daydreamer	Expresses interest in fantasy or daydreams (low score only if such interest is explicitly disavowed, middle score if not mentioned).	0.00	0.21	0.60
40	Guilt	Expresses guilt (about anything).	0.26	0.47	-0.09
41	Distance	Keeps partner at a distance, avoids development of any sort of interpersonal relationship (low	-0.27	0.33	0.33

		score implies behaviour to get close to the partner)			
42	Intellectual	Shows interest in intellectual or cognitive matters (e.g., by discussing an intellectual idea in detail or with enthusiasm).	0.58	0.25	0.04
43	Enjoy	Seems to enjoy the interaction.	0.87	-0.27	-0.07
44	Interesting	Says or does interesting things in this interaction.	0.77	-0.07	-0.06
45	Self-critical	Says negative things about self (e.g., is self-critical; expresses feelings of inadequacy).	0.48	0.53	-0.20
46	Ambitious	Displays ambition (e.g., passionate discussion of career plans, course grades, opportunities to make money).	0.50	0.23	-0.19
47	Blame	Blames others (for anything).	0.26	0.56	-0.25
48	Victim	Expresses self-pity or feelings of victimization.	0.20	0.58	-0.24
49	Attracted	Expresses sexual interest (e.g., acts attracted to partner; expresses interest in dating or sexual matters).	0.49	0.45	-0.09
50	Cheerful	Behaves in a cheerful manner.	0.74	-0.17	0.10
51	Givesup	Gives up when faced with obstacles (low score implies unusual persistence).	-0.39	0.40	0.34
52	Stereotypical	Behaves in a stereotypical masculine/feminine style or manner (apply the usual stereotypes appropriate to the subject's sex. Low score implies behaviour stereotypical of opposite sex)	0.14	0.34	0.29
53	Advisory	Offers advice.	0.45	0.56	-0.07
54	Fluent	Speaks fluently and expresses ideas well.	0.61	-0.16	-0.17
55	Accomplished	Emphasizes accomplishments of self, family, or housemates (low score = emphasizes failures of these individuals).	0.33	0.31	0.18
56	Competitive	Competes with partner(s) (low score implies cooperation).	0.23	0.35	0.56
57	Loud	Speaks in a loud voice.	0.52	-0.06	0.10
58	Sarcastic	Speaks sarcastically (e.g., says things (s)he does not mean; makes flippant comments that are not necessarily funny).	0.45	0.44	0.05
59	Physical	Makes or approaches physical contact with partner (of any sort, including standing unusually close without touching). (Low score implies unusual avoidance of physical contact, such as large interpersonal distance.)	0.42	0.18	0.51
60	Eyecontact	Engages in constant eye contact with partner. (Low score Implies unusual lack of eye contact.)	0.52	-0.13	-0.05
61	Detached	Seems detached from the interaction.	-0.40	0.58	-0.21
62	Quick	Speaks quickly (low score = speaks slowly).	0.20	0.13	0.26
63	Playful	Acts playful.	0.66	0.19	0.30
64	Advisor	Partner seeks advice from subject.	0.45	0.53	0.07
65	Head	Head is oriented towards partner during interaction	0.37	-0.25	-0.19
66	Body	Body is oriented towards partner during interaction	0.23	-0.34	-0.24

67	Pockets	Has hands in pockets during Interaction	-0.31	0.08	0.22
68	Props	Interacts with 'props' (i.e. phones, water bottles, pens)	0.02	0.49	-0.10
69	Window	Looks out of window	-0.17	0.36	0.19
70	Hair	Has fashionable hair (is following a trend)	0.09	0.52	0.08
71	Clothes	Has fashionable clothes (is following a trend)	0.25	0.08	0.00
72	Healthy	Looks generally healthy (a high score is notably healthy, low score is notably unhealthy, middle score is as would be expected)	0.19	0.02	-0.18
73	Fit	Looks physically fit (a high score is fitter than average, a low score is less fit than average, medium score is in sufficient fitness)	0.20	0.04	-0.18
74	Tidy	Touches/adjusts own clothes/jewellery	0.19	0.53	0.02
75	Groom	Touches/adjusts own body	0.23	0.39	0.16
76	Opening	Has a friendly opening gambit (a high score is overly friendly, middle is as would be expected, a low score is notably unfriendly)	0.74	-0.13	-0.01

Supplemental Figure.



Supplemental Figure. Scree plot to identify a three-factor solution for the behaviour coding dimension reduction, see Method section, Coding subsection.

Supplemental tables.

Supplemental Table A. *The estimates the effect of same sex dyads on the nature and accuracy of judgments at Time 1 (T1) with standard error of estimates (s.e.)*

Judgment	Target trait	T1 Accuracy	Same Sex Dyad	Interaction
Conscientiousness	Conscientiousness	-0.03 (0.31)	-0.01 (1.38)	0.07 (0.41)
Organised	Conscientiousness	-0.05 (0.35)	0.45 (1.54)	-0.11 (0.46)
Agreeableness	Agreeableness	-0.79 (0.37)*	-2.83 (1.87)	0.80 (0.51)
Friendly	Agreeableness	-0.60 (0.49)	-2.60 (2.45)	0.70 (0.67)
Neuroticism	Neuroticism	0.52 (0.23)*	2.06 (1.02)*	-0.66 (0.34)
Anxious	Neuroticism	0.55 (0.25)*	1.37 (1.11)	-0.44 (0.37)
Openness	Openness	-0.13 (0.28)	-3.08 (1.57)	1.00 (0.47)*
Creative	Openness	0.62 (0.25)*	-1.24 (1.33)	0.45 (0.40)
Extraversion	Extraversion	0.15 (0.21)	-0.36 (1.18)	0.18 (0.34)
Energetic	Extraversion	0.32 (0.23)	0.62 (1.25)	-0.13 (0.36)
Boldness	Boldness	-0.13 (0.38)	-1.82 (0.99)	1.05 (0.57)
Thrillseeking	Boldness	0.45 (0.40)	-0.62 (1.04)	0.54 (0.60)
Meanness	Meanness	0.03 (0.38)	-0.47 (0.57)	0.55 (0.58)
Friendly	Meanness	-0.46 (0.43)	-0.59 (0.64)	0.59 (0.64)
Disinhibition	Disinhibition	-0.04 (0.46)	-0.57 (0.69)	0.34 (0.66)
Organised	Disinhibition	-0.24 (0.44)	-0.82 (0.67)	0.95 (0.64)

Estimates are produced by a generalised liner model with the R formula: glm(Judgment ~Trait + SameSex + Trait*Samesex)
Same Sex Dyad is coded as 1= SameSex, 0= Not SameSex.
Uncorrected p values *p< .05, **p<.01, ***p< .001

Supplemental Table B. *The effect of same sex dyads (or not) on the nature and accuracy of judgments at Time 1 (T2)*

Judgment	Target trait	T2 Accuracy	Same Sex Dyad	Interaction
Conscientiousness	Conscientiousness	-0.01 (0.32)	-0.65 (1.41)	0.25 (0.42)
Organised	Conscientiousness	0.48 (0.38)	1.03 (1.69)	-0.22 (0.50)
Agreeableness	Agreeableness	-0.23 (0.36)	-3.13 (1.79)	0.90 (0.49)
Friendly	Agreeableness	0.26 (0.45)	0.11 (2.24)	0.05 (0.61)
Neuroticism	Neuroticism	0.98 (0.27)***	1.26 (1.20)	-0.47 (0.40)
Anxious	Neuroticism	0.57 (0.29)	0.05 (1.28)	-0.08 (0.42)
Openness	Openness	0.43 (0.28)	-0.91 (1.55)	0.33 (0.47)
Creative	Openness	0.56 (0.27)*	-0.19 (1.51)	0.20 (0.45)
Extraversion	Extraversion	0.54 (0.21)*	0.88 (1.16)	-0.14 (0.33)
Energetic	Extraversion	0.86 (0.22)***	3.27 (1.21)**	-0.92 (0.35)**
Boldness	Boldness	0.66 (0.42)	-0.48 (1.09)	0.36 (0.63)
Thrillseeking	Boldness	0.90 (0.40)*	0.51 (1.04)	-0.20 (0.60)
Meanness	Meanness	0.04 (0.31)	-0.31 (0.47)	0.42 (0.47)
Friendly	Meanness	-0.57 (0.39)	-0.04 (0.58)	0.33 (0.59)
Disinhibition	Disinhibition	0.71 (0.49)	0.62 (0.74)	-0.63 (0.71)
Organised	Disinhibition	-1.00 (0.48)*	-0.88 (0.73)	1.25 (0.70)

Estimates are produced by a generalised liner model with the R formula: glm(Judgment ~Trait + SameSex + Trait*Samesex)
Uncorrected p values *p< .05, **p<.01, ***p< .001