

This is the peer reviewed version of the following article: Sutherland, C. A., Rhodes, G., Burton, N. S. and Young, A. W. (2019), Do facial first impressions reflect a shared social reality?. Br J Psychol., which has been published in final form at <https://doi.org/10.1111/bjop.12390>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Do facial first impressions reflect a shared social reality?

Clare AM Sutherland^{1,2*}, Gillian Rhodes², Nichola S. Burton², Andrew W Young^{1,2}

¹ Department of Psychology, University of York, Heslington, York YO10 5DD, UK.

² School of Psychological Science, ARC Centre of Excellence in Cognition and its Disorders, University of Western Australia, Crawley 6009, WA, Australia.

* Corresponding author. Clare Sutherland, ARC Centre of Excellence in Cognition and its Disorders, School of Psychological Science, University of Western Australia, Crawley 6009, WA, Australia (email: clare.sutherland@uwa.edu.au).

Acknowledgements

This work was supported by an Australian Research Council (ARC) Discovery Early Career Research Award to CS [DE 190101043], the ARC Centre of Excellence in Cognition and its Disorders [CE110001021], and an ARC Discovery Award [DP170104602]. The funding sources had no influence on the research.

Declaration of interest: None.

Author contributions

FACIAL FIRST IMPRESSIONS

C. S. and A. Y. developed the study concept and study design. Data collection and analysis was performed by C. S. under the supervision of A. Young and with input from G. R. and N. B. C. S. drafted the manuscript, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

FACIAL FIRST IMPRESSIONS

Abstract

Influential facial impression models have repeatedly shown that trustworthiness, youthful-attractiveness and dominance dimensions subserve a wide variety of first impressions formed from strangers' faces, suggestive of a shared social reality. However, these models are built from impressions aggregated across observers. Critically, recent work has now shown substantial inter-observer differences in facial impressions, raising the important question of whether these dimensional models based on aggregated group data are meaningful at the individual-observer level. We addressed this question with a novel case series approach, using factor analyses of ratings of twelve different traits to build individual models of facial impressions for different observers. Strikingly, three dimensions of trustworthiness, youthful/attractiveness and competence/dominance appeared across the majority of these individual observer models, demonstrating that the dimensional approach is indeed meaningful at the individual level. Nonetheless, we also found differences in the stability of the competence/dominance dimension across observers. Taken together, results suggest that individual differences in impressions arise in the context of a largely common structure that supports a shared social reality.

Keywords: “first impressions”, “face perception”, “individual differences”

FACIAL FIRST IMPRESSIONS

Do facial first impressions reflect a shared social reality?

We meet strangers every day. From a glance at each face, we form an immediate impression of the person's character (Willis & Todorov, 2006). The validity of these impressions is still under debate, although there may be a small kernel of truth to some impressions (Bonnenfon, Hopfensitz, & De Neys, 2015; Olivola, Funk, & Todorov, 2014; Rhodes, Morley, & Simmons, 2013; Rule, Macrae, & Ambady, 2009). Regardless of the question of any underlying validity, however, it is clear that (valid or not) these facial impressions have important consequences across diverse social contexts (Olivola et al., 2014). For example, impressions of trustworthiness from profile images predict online financial lending rates (Duarte, Siegel, & Young, 2012), host popularity on Airbnb (Ert, Fleischer, & Magen, 2016), as well as sentencing outcomes in the criminal justice system (Wilson & Rule, 2015). In short, facial impressions clearly matter for society and the economy. Consequently, understanding facial impression formation has become a major research aim (see Stolier, Hehman, & Freeman, 2017; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015 for recent reviews).

A powerful approach to understanding the rich variety of facial impressions has been to use data-driven techniques to model key dimensions underlying these impressions (Oosterhof & Todorov, 2008). These techniques have revealed a surprising simplicity to the structure of facial impressions: two or three dimensions explain most of the variance across a wide range of facial impressions (Oosterhof & Todorov, 2008; South Palomares, Sutherland, & Young, 2018; Sutherland et al., 2013; Walker & Vetter, 2016). Most studies find trustworthiness and dominance (or competence) dimensions, while an additional youthful-attractiveness dimension also appears in some studies, largely where face photographs are more varied, particularly on age (South Palomares et al., 2018; Sutherland et al., 2013; although Wolfhechel et al., 2015 also found this dimension using standardised images). These dimensions link impressions to important forms of social appraisal involving threat

FACIAL FIRST IMPRESSIONS

avoidance and partner selection (Oosterhof & Todorov, 2008; South Palomares et al., 2018; Sutherland et al., 2013).

This dimensional approach represents a major advance because it shows that a tractable model can encompass a near-endless range of attributions; for example, Allport once famously noted that nearly 18,000 English words are devoted to trait impressions (Allport & Odbert, 1936). The dimensional models have been consequently highly influential and inspired substantial research, including studies investigating these dimensions across contexts, cultures and social stimuli (Stolier, Hehman, & Freeman, 2017; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015).

Interestingly, the dimensions of facial impressions mirror dimensions found generally across social cognition, even for non-visual stimuli. For example, Fiske's model of stereotypes of social groups has two dimensions, warmth and competence (Cuddy, Fiske, & Glick, 2008; Fiske, Cuddy, & Glick, 2007), which have been shown to align with trustworthiness and dominance dimensions, respectively (Sutherland, Oldmeadow, & Young, 2016; Walker & Vetter, 2016). Rosenberg's model of personality impressions of abstract targets has similar social and intelligence dimensions (Rosenberg, Nelson, & Vivekananthan, 1968). Fletcher and colleagues' (2000) model of partner preferences has three dimensions of warmth, vitality and status, which highly correspond with trustworthiness, youthful-attractiveness and dominance (South Palomares et al., 2018; three similar dimensions also appear in Osgood's 1969 semantic differential model of attitudes, and in Rosenberg et al. 1968). At the same time, the facial impression dimensions can be recovered from mapping visual cues present in face images themselves, showing that there is an intimate link between these impressions and visual information (Vernon, Sutherland, Young, & Hartley, 2014). It is striking that a relatively simple model can span impressions made from a split-second glimpse at a face image, through to more complex conceptual attitudes.

FACIAL FIRST IMPRESSIONS

However, there is an issue with the approach employed in studies of face perception so far. Critically, dimensional models are theorised to describe facial impression formation as it occurs in individual perception. Yet, impressions are modelled at the level of the faces, after averaging across individual observers' impressions (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2016). The models may not, therefore, capture the structure of impressions for *individual* people. This issue is critical for the theory behind the models, because it assumes that these dimensions structure a common social reality, shared across individual observers.

Why have facial impression models been built on group data?

The use of group-level data has been driven by the observation that participants substantially agree in their impressions, suggesting that facial impressions reflect a shared social reality (Walker & Vetter, 2016; also Sutherland et al., 2013; Todorov et al., 2015). But *do* individual observers agree in their impressions?

The claim that observers largely agree in their impressions is often based on the finding of high values of Cronbach's alpha, where alpha is calculated across individual observers for a given judgement and set of faces, as if observers were items in a test (for a review, see Todorov, Said, & Verosky, 2011). This use of alpha is appropriate when drawing conclusions at the group level, because alpha measures the extent to which the *group* of items (here, observers) will agree with other groups tested in future (Cortina, 1993), but it is inappropriate for measuring agreement at the individual level because alpha is inflated by the number of items (i.e. observers: Cortina, 1993). Thus, studies citing high alpha values do not necessarily demonstrate that individual observers agree in their impressions.

Better evidence for agreement in impressions across observers comes from three sources. First, Zebrowitz and her colleagues have measured the degree to which individual observers agree with members of the same or different social groups, for example, finding

FACIAL FIRST IMPRESSIONS

that older and younger people do agree in their impressions of faces across age (Zebrowitz, Franklin Jr, Hillman, & Boc, 2013). Second, studies modelling group-based impressions also usually cross-validate these impression dimensions by constructing computer-generated (Oosterhof & Todorov, 2008; Walker & Vetter, 2016) or morphed (Sutherland et al., 2018, 2013) faces directly from the models, and then ascertaining that these faces are perceived as expected by new sets of observers. Third, the fact that facial impressions predict important social consequences in the real world is also suggestive that impressions must be consensual, at least to some extent.

So, there are reasons to expect some degree of underlying consensual agreement between the impressions of different observers. Strikingly, however, a recent study has shown greater scope for observer disagreement in facial impressions than hitherto recognised (Hehman, Sutherland, Flake, & Slepian, 2017). Using a multilevel modelling approach, the study showed that around half of the variation in perceived impressions of trustworthiness, dominance, and attractiveness was due to idiosyncratic preferences that differ across observers, not consensus impressions (Hehman et al., 2017; see also Germine et al., 2015; Hönekopp, 2006). This study is important because it focuses attention on the extent of individual differences in facial impressions across observers, building on a long tradition of social psychological work which examines consensus in impressions outwith face perception (e.g. Kenny, 1991). This new study also raises the question as to what extent the structure of facial impression dimensions may also differ across single observers and from single observers to groups.

To summarise, existing studies have shown that consensus impressions (and resulting dimensions) are stable across groups of participants. Whilst studies have also showed agreement in facial impressions across individual or groups of observers (e.g. Zebrowitz et al., 2013), there may be more room for individual differences than previously understood (Hehman et al., 2017). Importantly, what has yet to be examined is the theoretical issue of

FACIAL FIRST IMPRESSIONS

whether the consensual dimensions of facial impressions are in fact the same for individual observers. This question has yet to be tested.

Current study

We tested whether consensual dimensions of facial impressions also subserved impressions made by individual observers. Answering this question is critical because models of facial impressions presume to describe the structure of impressions in individual perception, not as a social construct that only emerges at a group level (Sutherland et al., 2013; Todorov et al., 2015). It is also timely, because understanding what drives stable individual differences in facial impressions has become a major new research direction (Germine et al., 2015). Without establishing that these dimensions hold across observers, individual difference studies (e.g. Hehman et al., 2017), risk missing more fundamental variation in the basic structure of impressions.

To test whether dimensional models apply at the individual level, we applied a novel 'case series' approach, using factor analyses of ratings of twelve different traits to build models of facial impressions for individual observers. Our use of a case series approach was inspired by research in cognitive neuropsychology, where similar concerns about the relationship between data involving individuals and overall group means applies (Schwartz & Dell, 2010). We reasoned that if theoretical models derived from group performance are valid at the individual level, then the three dimensions of trustworthiness, attractiveness and dominance should emerge for most observers. This finding would provide novel evidence that these dimensions are meaningful at the observer level, despite any idiosyncratic variation in impressions.

Importantly, our approach was highly data-driven. The face images that participants rated were everyday photographs of Caucasian adult faces taken from the internet, with all the variability in pose, expression, age, hairstyle, lighting and camera properties that entails;

FACIAL FIRST IMPRESSIONS

nothing was standardised. The twelve traits participants were asked to rate were chosen to represent traits from previous group-based studies; these included characteristics that people use spontaneously when they describe faces (Oosterhof & Todorov, 2008; Sutherland et al., 2018). In these ways we sought to encompass the range of spontaneous evaluations of everyday images of faces without imposing a predetermined dimensional structure or even a fixed number of underlying dimensions. Instead, the structure of each participant's impressions was determined solely on the basis of their pattern of responses.

Method

Simulations

We first established the validity of our intuition that the structure of facial impressions *can* differ between individual and group-based data. To achieve this demonstration we simulated the simplest possible case based on a variant of Oosterhof and Todorov's (2008) seminal two-dimensional model of group data. We were interested to know whether it is possible, in principle at least, that an overall two-dimensional model might emerge simply from combining data from different individuals, each of whom used only one of the two putative dimensions.

We therefore created a small-scale simulation based around hypothetical sets of data for two observers rating 36 imaginary 'faces' on three traits for each of two putative dimensions (six traits in total). Each of these two hypothetical observers had *one* dimension (corresponding either to 'trustworthiness' or to 'dominance') underlying their impressions on the traits. By randomly jittering the individual face ratings of these hypothetical 'one-dimensional' observers we then created multiple hypothetical one-dimensional data sets. These data sets were then pooled and submitted to individual-based or group-based PCA to establish whether the solutions might differ in terms of the numbers of factors identified. These simulated data showed that a two-dimensional trustworthiness by dominance model

FACIAL FIRST IMPRESSIONS

can emerge at the group level from the responses of hypothetical observers who individually only show one of these dimensions (Table S1) or who (because of the random variation introduced into the hypothetical data) exhibit what appear to be more complex individual dimensional structures (Table S2). Importantly, in our simulations, the overall group-based model did not correspond to *any* of the individual observers' impressions, justifying the need for an empirical test with real observers.

Participants

Twenty-four British participants (12 male: mean age: 20.6 years, SD age: 2.5 years) volunteered to take part in return for course credit or small remuneration. Participants were Caucasian, culturally Western and provided informed consent to procedures that were approved by the ethics committee of the University of York Psychology Department. Two additional participants were tested but dropped out; their data were not analysed further. Our sample size was chosen based on the in-depth case series approach used, and so that two participants (one male) each completed one of twelve counterbalancing orders (see Design).

Stimuli

A set of two hundred images of unfamiliar faces (100 female) were chosen at random from a larger image dataset containing 1,000 face images collected from the internet and used in previous work (Santos & Young, 2005, 2008, 2011; Sutherland et al., 2013). Naturalistic images were used to capture the variation in potential social cues present when observing others in everyday life or when browsing online (Jenkins et al., 2011; Sutherland, Young, & Rhodes, 2016 discuss the importance of using naturalistic face images).

Design

Participants rated the set of 200 faces on twelve different traits, each in a separate block with

FACIAL FIRST IMPRESSIONS

24 practice faces presented before each new block (practice faces were not analysed). In order to be able to build individual models of impressions for each observer separately, it was essential that all observers rated a set of faces on all traits. In order to remove the confounding effect of trait order, the order of trait blocks was counterbalanced across participants using a Latin square design. The twelve traits were chosen for theoretical reasons to test the dimensions described by previous work. That is, if the dimensions do exist at the individual observer level, we should reasonably expect to find them represented in this collection of traits. Approachability, trustworthiness, smiling and warmth were selected to index the trustworthiness dimension identified by previous work (Oosterhof & Todorov, 2008; Sutherland et al., 2013); dominance, competence, intelligence and masculinity (feminine to masculine) were chosen to index the dominance dimension (Oosterhof & Todorov, 2008; Sutherland et al., 2013); and age, babyfacedness, health, and attractiveness were chosen to index the youthful-attractiveness dimension (Sutherland et al., 2013; Wolffhechel et al., 2015). Data collection was split into two sessions, each less than an hour, to avoid fatiguing participants. Group-level ratings of trustworthiness, dominance, warmth and competence were also used in a different study with a separate aim (Sutherland, Oldmeadow, et al., 2016).

On each trial, participants saw one of the 200 images, randomly selected. They were then asked to rate that face for their impression on a scale from 1 to 7, ranging from no smile-big smile, or (very) unapproachable–approachable, untrustworthy–trustworthy, cold–warm, nondominant–dominant, feminine–masculine, incompetent–competent, unintelligent–intelligent, young adult–old adult, maturefaced–babyfaced, unhealthy–healthy, and unattractive–attractive. Trials were self-paced with an inter-stimulus interval of 750ms. Participants were asked to go with their gut instinct and not to take too long on any individual face. On average, participants took 2.1 seconds to respond to each face.

Data

FACIAL FIRST IMPRESSIONS

The raw data is uploaded along with the paper.

Results

Individual factor analyses

Factor labelling criteria: In order to understand the dimensionality of individual participants' impressions, we first entered the twelve impression judgements into separate factor analyses for each participant. We used the same method that previous authors have used when modelling impressions at the group level (e.g. Sutherland et al., 2018, 2013). As in previous work, the number of factors reported was determined independently by Kaiser's criterion, the scree test, parallel analysis and MAP analyses (see Fabrigar, Wegener, MacCallum, & Strahan, 1999; Kline, 1994). The final number of rotated factors was based on the modal result across these four tests, with Kaiser's criterion used as a default where there was no mode factor number. **Table 1** summarises the outcomes of the dimensionality tests and overall dimensions emerging for each individual participant, and **Table 2** depicts the average factor loadings on each dimension (**Figure S1** depicts the individual scree tests and **Table S3** depicts the individual factor analyses).

A factor was labelled *trustworthiness* if the highest loading on that factor was from warmth, trustworthiness, approachability or smiling, and at least two of these traits loaded on that factor at .3 or higher. Similarly, a factor was labelled *youthful/attractiveness* if the highest loading was from age, babyfacedness, attractiveness or health, and at least two of these attributes loaded on this factor at .3 or higher. A factor was labelled *competence/dominance* if the highest loading was from competence, intelligence, dominance or masculinity, and at least two of these attributes loaded at .3 or higher. For all dimensions, we also allowed high loadings from age or masculinity, as these social category cues often relate to multiple dimensions (Sutherland et al., 2013). Interestingly, we also observed a split between competence and dominance so we also report the results with competence and dominance as

FACIAL FIRST IMPRESSIONS

separate dimensions. Any other dimension emerging was labelled with the highest loading on that factor.

Trustworthiness, youthful/attractiveness and competence dimensions appear at the individual observer level

In order to test whether the predicted dimensions emerged, we coded the presence or absence of a predicted dimension for each participant using the criteria above. We then statistically tested whether these dimensions were present in the majority of participants using two-tailed binomial tests.

Examining each dimension separately, a significant majority of the participants (96%) showed a trustworthiness dimension as expected (23 people, i.e. all but one person; two-tailed binomial test: $p < .001$). A significant majority of the participants also showed a youthful-attractiveness dimension, which was also very robust across participants (92% of the participants, 22 people, $p < .001$; the remaining two participants had a ‘youth’ only dimension, without a high loading from attractiveness). Also in line with our predictions, a significant majority of the participants (83%) also showed a competence or dominance dimension (20 people, $p = .002$; **Table 1**). Only 17% of the participants showed a unique, unpredicted dimension (4 people; a significant minority, $p = .002$; **Table 1**).

Table 1 around here

Overall, a significant majority of the participants showed all three dimensions of trustworthiness, youthful/attractiveness, and competence/dominance (75%; 18 out of 24 people; $p = .023$; **Table 1**; of these, nine people showed a split between dominance and competence). Two further participants showed a youth-only dimension (increasing the majority to 79%; $p = .002$; **Table 1**; one of whom showed the predicted three dimensions if

FACIAL FIRST IMPRESSIONS

this dimension was included). The remaining six participants all showed two out of the three predicted dimensions. Taken together, this pattern provides good evidence that the dimensions emerge at the individual participant level.

Interestingly, although we had initially assumed that competence and dominance would form one dimension, competence and dominance instead appeared to split for half of the participants (**Table 1**; see also Sutherland et al. 2016). That is, eleven participants showed a combined competence/dominance dimension, and a further nine either had a competence-only dimension (six people, including intelligence), a dominance-only dimension (one person) or both dimensions (two people). Moreover, masculinity did not appear to contribute greatly to this dimension when it was expressed as competence, against predictions (see **Table 2**). Instead, masculinity may be more related to dominance, and the use of this cue appeared to differ across participants (returned to in the **Discussion**).

Table 2 around here

Mapping to consensus dimensions

Given that individual observers showed a similar structure to facial impressions to that derived from theories built on group consensus data, one key question is whether individual observers also use similar facial cues to form these impressions, as would be predicted by consensus models. We addressed this question by correlating each participant's individual factor scores with independently derived factor scores for the same faces, taken from a previous, group-level model of facial impressions (Sutherland et al., 2013; **Figure 1A** depicts these correlations). If observers in the current study are basing their impressions on similar facial cues, as predicted by consensus models, then correlations between predicted similar dimensions should be significant and high because they are based on the same set of faces. As labelling of factors can be subjective, this analysis also acts as a formal quantification of the

FACIAL FIRST IMPRESSIONS

appropriateness of our choice of factor labels. All statistical tests were run after Fisher transformation to account for the bounded nature of correlation coefficients.

Importantly, as predicted, the currently obtained trustworthiness scores correlated highly with the previously obtained trustworthiness scores (average $r = .79$, $t(22) = 24.11$, $p < .001$, statistically tested by comparing the participants' correlations to zero with a one-sample t -test; **Figure 1B**). The current youthful-attractiveness scores also highly correlated with the previously obtained youthful-attractiveness scores (average $r = .74$, $t(23) = 23.91$, $p < .001$; **Figure 1B**). The current competence/dominance factor correlated moderately with the previously obtained dominance scores (average $r = .40$, $t(19) = 12.51$, $p < .001$; **Figure 1B**). However, the competence/dominance correlations were less strong than the trustworthiness: $t(18) = 12.00$, $p < .001$, $d = 3.49$; or youthful-attractiveness correlations: $t(19) = 13.31$, $p < .001$, $d = 2.82$. The trustworthiness correlations were also slightly but significantly stronger than the youthful-attractiveness correlations: $t(22) = 2.33$, $p = .029$, $d = 0.50$. This pattern indicates that observers were largely basing their impressions on highly similar facial cues to those predicted by consensus models for trustworthiness and youthful-attractiveness, with the greatest individual differences found for competence/dominance.

As a secondary test of the comparability between individual and consensus models, we entered these correlations into a 3 (current participants' dimension scores) x 3 (previously-obtained dimension scores) repeated measures ANOVA (**Figure 1B**; see **Table S2** for individual values). As expected, we found a significant interaction, so that the highest correlations were between the predicted congruent dimensions from the previous model: $F(4,72) = 198.37$, $p < .001$, $\eta^2 = .92$. That is, each individual dimension found here corresponded most highly with the congruent dimension taken from the previously-obtained model than with the other two dimensions in this previous model (trustworthiness: $t(22) \geq 18.03$, $p \leq .001$, $d \geq 7.16$, youthful-attractiveness: $t(23) \geq 11.22$, $p \leq .001$, $d \geq 3.74$; and competence/dominance: $t(19) \geq 3.44$, $p \leq .005$, $d \geq 1.25$ (**Figure 1B**). Thus, the individual

FACIAL FIRST IMPRESSIONS

differences in the competence/dominance dimension appear to be specific to this dimension, rather than affecting the entire structure of the model.

Figure 1 around here

Relative ordering of dimensions

We then tested if the dimensions emerged in a meaningful order, with the prediction that the trustworthiness dimension should explain most of the variance across individual observer models. We based this prediction on the general theory that trustworthiness (or warmth, a closely related trait) is the primary dimension of social appraisal (Cuddy et al., 2008). If so, then the trustworthiness dimension should in general emerge first across observers, so that it explains the majority of the variance. To test this prediction, we used χ^2 tests to compare the proportion of participants with trustworthiness, youthful/attractiveness, competence/dominance or other dimensions at each of the first five factor positions (see **Figure 2**). Here the null hypothesis represents an even split of these four dimension types in each position across participants, so that the dimension type is not predictive of variance explained (note that this null is a conservative test of our predictions, as the ‘other’ category is a catch-all category for *any* other result). All tests with $df = 1$ were conservatively corrected for continuity (Yates, 1934).

Figure 2 around here

Overall, the first dimension was distributed significantly differently from chance: $\chi^2(3) = 38, p < .001$; as predicted, the first dimension was significantly more likely to be trustworthiness: $\chi^2(1) = 37.69, p < .001$ (**Figure 2**). The second dimension was also distributed significantly differently from chance: $\chi^2(3) = 14.33, p < .002$, and more likely to

FACIAL FIRST IMPRESSIONS

be youthful-attractiveness: $X^2(1) = 6.72, p < .010$ (**Figure 2**). The third dimension was again distributed significantly differently from chance: $X^2(3) = 9.87, p = .020$, and marginally more likely to be the competence/dominance dimension: $X^2(1) = 3.27, p = .071$. Distribution of the fourth and fifth dimensions did not significantly differ (both chi2: $X^2(3) < 3.67, p > 0.29$). In summary, as predicted, trustworthiness explained the most variance across individual participant models, followed by youthful-attractiveness and then competence/dominance.

Trait correlations at the individual level

Finally, **Figure 3A** displays the relationships between the twelve different trait impressions for each participant separately. A strikingly similar pattern is clearly shown across different participants' impressions (**Figure 3A**). This test formally quantifies the similarity of the structure of impressions across observers, demonstrating that our principal conclusion of substantial inter-observer consistency in the underlying structure of trait evaluations does not depend on the specifics of the factor analyses chosen.

We statistically verified the high similarity across participants by correlating the pattern of trait relationships across participants (see **Figure 3B**). On average, the participants showed high similarity in the pattern of their impressions, significantly higher than zero: average $r = .77, t(23) = 41.28, p < .001$ (statistical testing after Fisher transformation to correct for bounded correlation coefficients).

Discussion

We tested whether theoretical models of facial impressions derived from group-based responses also apply to impressions formed by individual observers. To achieve this aim, we adopted a data-driven approach that involved modelling the structure of facial impressions of highly variable everyday images of faces as made by individual observers. In effect, we treated the data from each participant's ratings across twelve different traits as a separate

FACIAL FIRST IMPRESSIONS

individual case study, and then evaluated the consistency of patterns of responses across the entire case series. To the best of our knowledge this study is the first to employ this essential test of facial impression models by data-driven modelling of the structure of individual participant responses.

Importantly, we found that dimensions of trustworthiness, youthful-attractiveness and competence/dominance did explain the impressions made by most observers. This finding suggests that leading models of facial impressions (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2009) do reflect a common structure across individual observers. Of course, this conclusion can at present only apply to the young, culturally Western student population we tested. However, this is also the population for which existing theories have been primarily created and our findings show that for this population the individual dimensions are strikingly consistent. This result is critical because currently the field interprets these models as if they describe common dimensions of impressions used by individual observers, yet the dimensions have previously been entirely derived from aggregated data, which need not accurately reflect any individual observer's impressions (see simulated data, Tables S1-S2). Our results validate the widely used interpretation that group-level data can reflect a shared social reality and demonstrate that future research can meaningfully focus on these same broadly shared dimensions when seeking to understand individual differences in facial impressions (cf. Hehman et al., 2017).

We found that trustworthiness explained the most variance across observers, followed by youthful-attractiveness and then competence/dominance. Trustworthiness (or the related traits of approachability or warmth) has previously been theorised to be the most important dimension of impressions (Abele & Bruckmüller, 2011; Cuddy et al., 2008). Our findings support the primacy of trustworthiness for facial impressions made by individual observers.

The shared social reality of facial impressions

FACIAL FIRST IMPRESSIONS

Our findings provide strong support for the theory that the structure of facial impressions largely reflects a shared social reality. Following prominent models (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Walker & Vetter, 2016), we suggest that the trustworthiness dimension represents the judgement of others' intentions (good or bad), the competence or dominance dimension represents the judgement of others' ability to carry out their intentions, and the youthful-attractiveness dimension represents judgement of potential mate value involved in sexual selection.

Why are these particular dimensions so important? Likely, impressions along these dimensions are critical for a number of reasons, including adaptive pressures to judge conspecifics for threat and potential mate value (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Zebrowitz, 2017; Zebrowitz & Montepare, 2014). For example, one adaptive cue to threat is emotional expression (Darwin, 1872), which forms an important facial cue to trustworthiness (Sutherland, Young, et al., 2016, although not the only cue: Vernon, Sutherland, Young, & Hartley, 2014). Overgeneralisation of adaptive responses to these kinds of facial cues has led to facial impressions that are often misleading in modern contexts (Olivola et al., 2014). Facial impressions are therefore akin to stereotypes, albeit prompted by visual cues (Oldmeadow, Sutherland, & Young, 2013).

We are not seeking to imply that facial impressions are *only* driven by bottom-up visual cues. As reviewed in the introduction, there are close links between face perception models and models across social cognition, including models of stereotyping, partner preferences and intergroup prejudice (South Palomares et al., 2018; Sutherland, Oldmeadow, et al., 2016; Walker & Vetter, 2016). Thus, the dimensions of first impressions from faces may reflect the way people think about all manner of stimuli, itself suggesting that processing of facial impressions from visual cues is deeply entwined with top-down conceptual attitudes.

Observer variation in facial impressions

FACIAL FIRST IMPRESSIONS

Our findings do not rule out individual differences in impression formation. Instead, they help understand where and why individual differences are to be found. Whilst there is now convincing evidence that observers can substantially differ in their facial impressions (Germine et al., 2015; Hehman et al., 2017; Hönekopp, 2006), our findings demonstrate an underlying commonality in the structure of facial impressions. Critically, individual differences in impressions seem to arise as variations within this common structure, and emerge more strongly for the less salient dimensions.

An important question for future research involves disentangling systematic individual differences from differences in reliability. That is, participants who show the clearest factor structure may in part be those who attend most carefully to the task and give the most consistent responses, as well as those who show more fundamental individual differences in facial impressions. In this respect, we did find evidence of striking observer differences, suggestive of systematic idiosyncrasy. Specifically, the competence/dominance dimension was far more flexible across observers than trustworthiness and youthful-attractiveness. This result is consistent with Hehman and colleagues' (2017) conclusion that the dominance dimension is particularly idiosyncratic, over and above differences in observer reliability. It also agrees with the pattern found across culture: for example, Sutherland et al. (2018) found similar impression dimensions for British and Chinese groups, but with the least cross-cultural agreement on competence. Moreover, Sutherland et al. (2016) failed to find an overall dominance factor; similar to our finding that competence was more robust than dominance across observers. Finally, Zebrowitz and colleagues (2012) also found more cross-cultural agreement on warmth-related facial impressions as compared to competence-related impressions.

Possibly, there is more agreement for factors that include more directly observable facial qualities. Age and attractiveness are clearly visible in the face (Rhodes, 2006), and trustworthiness is strongly influenced by emotional expression (Sutherland, Young, et al.,

FACIAL FIRST IMPRESSIONS

2016), whereas dominance or (especially) competence may simply be less obviously visual (also argued by (Hehman et al., 2017).

Alternatively, and not mutually exclusively, we suggest that dominance (or competence) is the most variable dimension because these judgements are highly contextual: competence naturally reflects ability at a *particular* task, while dominance reflects power over a *particular* situation or group (see also Sutherland, Oldmeadow, et al., 2016). Here, impressions were made without a defined context, following the procedure used by Oosterhof and Todorov (2008) and in most other studies. Thus, depending on the context they are drawing on to make these judgements, different people may be using different facial cues to judge dominance or competence; for example, masculinity versus health, as suggested by individual differences in factor loadings. Indeed, masculinity deviated most strongly from predictions based on consensus models. These suggestions lead to testable predictions around which individual differences influence facial impression formation. For example, people who strongly endorse traditional gender stereotypes may be particularly influenced by facial masculinity when judging competence or dominance. Importantly, the suggestion that people vary in impression formation is highly compatible with the overarching theory that selection pressures have shaped impression formation (Oosterhof & Todorov, 2008), as selection pressure itself relies on variation (Penton-Voak & Morrison, 2011).

Future directions

We focus here on faces because dimensional models have been especially influential in understanding face perception (Todorov et al., 2015). However, the dimensional modelling approach has been increasingly influential in explaining other aspects of social perception. Our current approach can be applied in future to test dimensional models of social perception across fields, including for voices (McAlear, Todorov, & Belin, 2014), bodies (Morrison, Wang, Hahn, Jones, & DeBruine, 2017) and relationship partner preferences (South

FACIAL FIRST IMPRESSIONS

Palomares et al., 2018). Our approach can also be extended to test other populations across age and culture.

It is also crucial that the field addresses the contributions of different mechanisms underlying systematic observer variation in facial impressions. Potential mechanisms include differences in beliefs, including endorsement of gender stereotypes (Eagly & Steffen, 1984) or cultural beliefs (Sutherland et al., 2018), social or perceptual experience (Dotsch, Hassin, & Todorov, 2016; Verosky & Todorov, 2013), motivation (Jones, Little, Watkins, Welling, & DeBruine, 2011), personality (Matarozzi, Todorov, Marzocchi, Vicari, & Russo, 2015) or inherent differences in visual perception (Watkins et al., 2010). We have shown that the dimensions themselves are largely robust across observers, making it possible to use impressions along these dimensions as a starting point for this investigation.

Conclusions

We build the first data-driven individual observer models of facial impressions to create a strong test of whether influential models of facial impression formation reflect a shared social reality. Across observers, we find high similarity in facial impression dimensions, supporting current theoretical models. However, we also find striking differences, with the dominance or competence dimension being particularly flexible across observers. Our results provide a starting point for research to understand individual differences in facial impressions and suggest that individual differences in impressions may mainly occur as deviations within a common overall structure.

References

Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the “Big Two”? Preferential processing of communal information. *Journal of Experimental Social Psychology, 47*, 935–948. <https://doi.org/10.1016/j.jesp.2011.03.028>

FACIAL FIRST IMPRESSIONS

- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1), 1–171. <http://dx.doi.org/10.1037/h0093360>
- Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in Cognitive Sciences*, 19(8), 421–422. <https://doi.org/10.1016/j.tics.2015.05.002>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, 40, 61–149. [https://doi.org/10.1016/S0065-2601\(07\)00002-0](https://doi.org/10.1016/S0065-2601(07)00002-0)
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. Retrieved April 24, 2018, from <https://www.gutenberg.org/files/1227/1227-h/1227-h.htm>
- Dotsch, R., Hassin, R. R., & Todorov, A. T. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour*, 1, 1–6. <https://doi.org/10.1038/s41562-016-0001>
- Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies*, 25(8), 2455–2484. <https://doi.org/10.1093/rfs/hhs071>
- Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, 46(4), 735–754. <http://dx.doi.org/10.1037/0022-3514.46.4.735>
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos on Airbnb. *Tourism Management*, 55, 62–73. <http://dx.doi.org/10.1016/j.tourman.2016.01.013>

FACIAL FIRST IMPRESSIONS

- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <http://dx.doi.org/10.1037/1082-989X.4.3.272>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000). Ideals, perceptions, and evaluations in early relationship development. *Journal of Personality and Social Psychology*, 79(6), 933–940.
- Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A. M., Smoller, J. W., Kwok, H., ... Wilmer, J. B. (2015). Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Current Biology*, 25(20), 2684–2689. <https://doi.org/10.1016/j.cub.2015.08.048>
- Hehman, E. A., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *JPSP*, 113(4), 513–529. <http://dx.doi.org/10.1037/pspa0000090>
- Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 199–209. <https://doi.org/10.1037/0096-1523.32.2.199>
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121, 313–323. <https://doi.org/10.1016/j.cognition.2011.08.001>
- Jones, B. C., Little, A. C., Watkins, C. D., Welling, L. L. M., & DeBruine, L. M. (2011). Reported sexual desire predicts men's preferences for sexually dimorphic cues in women's faces. *Archives of Sexual Behavior*. <https://doi.org/10.1007/s10508-010-9721-1>

FACIAL FIRST IMPRESSIONS

- Kenny, D. A. (1991). A General Model of Consensus and Accuracy in Interpersonal Perception. *Psychological Review*, 98(2), 155–163. <http://dx.doi.org/10.1037/0033-295X.98.2.155>
- Kline, P. (1994). *An easy guide to factor analysis*. London; New York: Routledge.
- Mattarozzi, K., Todorov, A., Marzocchi, M., Vicari, A., & Russo, P. M. (2015). Effects of gender and personality on first impressions. *PLoS ONE*, 10(9), e0135529. <https://doi.org/10.1371/journal.pone.0135529>
- McAler, P., Todorov, A., & Belin, P. (2014). How do you say ‘Hello’? Personality impressions from brief novel voices. *PLoS ONE*, 9(3), e90779. <https://doi.org/10.1371/journal.pone.0090779>
- 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. <https://doi.org/10.1371/journal.pone.0185093>
- Oldmeadow, J. A., Sutherland, C. A. M., & Young, A. W. (2013). Facial stereotype visualization through image averaging. *Social Psychological and Personality Science*, 4(5), 615–623. <https://doi.org/10.1177/1948550612469820>
- Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, 18(11), 566–570. <https://doi.org/doi:10.1016/j.tics.2014.09.007>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *PNAS*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Osgood, C. E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, 12(3), 194–199.
- Penton-Voak, I. S., & Morrison, E. R. (2011). Structure, expression, and motion in facial attractiveness. In A. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *Oxford Handbook of Face Perception*. OUP Oxford.

FACIAL FIRST IMPRESSIONS

- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 57(1), 199–226.
<https://doi.org/10.1146/annurev.psych.57.102904.190208>
- Rhodes, G., Morley, G., & Simmons, L. W. (2013). Women can judge sexual unfaithfulness from unfamiliar men's faces. *Biology Letters*, 9(1), 20120908.
<https://doi.org/10.1098/rsbl.2012.0908>
- Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, 9(4), 283–294. <http://dx.doi.org/10.1037/h0026086>
- Rule, N. O., Macrae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science*, 20(4), 441–443.
<https://doi.org/10.1111/j.1467-9280.2009.02314.x>
- Santos, I. M., & Young, A. W. (2005). Exploring the perception of social characteristics in faces using the isolation effect. *Visual Cognition*, 12(1), 213–247.
<https://doi.org/10.1080/13506280444000102>
- Santos, I. M., & Young, A. W. (2008). Effects of inversion and negation on social inferences from faces. *Perception*, 37(7), 1061–1078. <https://doi.org/10.1068/p5278>
- Santos, I. M., & Young, A. W. (2011). Inferring social attributes from different face regions: Evidence for holistic processing. *The Quarterly Journal of Experimental Psychology*, 64(4), 751–766. <https://doi.org/doi/full/10.1080/17470218.2010.519779>
- Schwartz, M. F., & Dell, G. S. (2010). Case series investigations in cognitive neuropsychology. *Cognitive Neuropsychology*, 47, 477–494.
<https://doi.org/10.1080/02643294.2011.574111>
- South Palomares, J. K., Sutherland, C. A. M., & Young, A. W. (2018). Partner preference and facial first impressions models: Comparable or distinct underlying structures? *British Journal of Psychology*, 109(3), 538–563. <https://doi.org/doi:10.1111/bjop.12286>

FACIAL FIRST IMPRESSIONS

- Stolier, R. M., Hehman, E., & Freeman, J. B. (2017). A dynamic structure of social trait space. *Trends in Cognitive Sciences*, 22(3), 197–200.
<https://doi.org/10.1016/j.tics.2017.12.003>
- 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). <https://doi.org/10.1177/0146167217744194>
- Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127(1), 105–118.
<https://doi.org/10.1016/j.cognition.2012.12.001>
- Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, 157, 257–267. <https://doi.org/10.1016/j.cognition.2016.09.006>
- Sutherland, C. A. M., Young, A. W., & Rhodes, G. (2016). Facial first impressions from another angle: Social judgements are influenced by both emotional expression and viewpoint. *British Journal of Psychology*, 108(2), 397–415.
<https://doi.org/10.1111/bjop.12206>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, 66(1), 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., Said, C. P., & Verosky, S. C. (2011). Personality impressions from facial appearance. In A. Calder, G. Rhodes, M. Johnson, & J. V. Haxby (Eds.), *Oxford Handbook of Face Perception*. (pp. 630–651). Oxford: Oxford University Press.

FACIAL FIRST IMPRESSIONS

- Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *PNAS*, *111*(32), E3353–E3361. <https://doi.org/10.1073/pnas.1409860111>
- Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology*, *49*(4), 661–669. <https://doi.org/10.1016/j.jesp.2013.02.004>
- 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. <https://doi.org/10.1167/9.11.12>
- Walker, M., & Vetter, T. (2016). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, *110*(4), 609–624. <https://doi.org/10.1037/pspp0000064>
- Watkins, C. D., Fraccaro, P. J., Smith, F. G., Vukovic, J., Feinberg, D. R., DeBruine, L. M., & Jones, B. C. (2010). Taller men are less sensitive to cues of dominance in other men. *Behavioral Ecology*, *21*(5), 943–947. <https://doi.org/10.1016/j.paid.2010.08.006>
- 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. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331. <https://doi.org/10.1177/0956797615590992>
- Wolffhechel, K., Hahn, A. C., Jarmer, H., Fisher, C. I., Jones, B. C., & DeBruine, L. M. (2015). Testing the utility of a data-driven approach for assessing BMI from face images. *PLoS ONE*, *10*(10), 1–10. <https://doi.org/10.1371/journal.pone.0140347>

FACIAL FIRST IMPRESSIONS

Yates, F. (1934). Contingency table involving small numbers and the χ^2 test, *1*(2), 217–235.

<https://doi.org/10.2307/2983604>

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological*

Science, *26*, 237–242. <https://doi.org/10.1037/a0030927>

Zebrowitz, L. A., Franklin Jr, R. G., Hillman, S., & Boc, H. (2013). Older and younger

adults' first impressions from faces: Similar in agreement but different in positivity.

Psychology and Aging, *28*(1), 202–212. <http://dx.doi.org/10.1037/a0030927>

Zebrowitz, L. A., & Montepare, J. M. (2014). Faces and first impressions. In J. A. Bargh & G.

Borgida (Eds.), *Handbook of Personality and Social Psychology, Vol. 1: Attitudes and*

Social Cognition. (Vol. 1). Washington D.C.: American Psychological Association.

Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V.,

& Godoy, R. (2012). First impressions from faces among US and culturally isolated

Tsimane' people in the Bolivian rainforest. *Journal of Cross-Cultural Psychology*,

43(1), 119–134. <https://doi.org/10.1177/0022022111411386>

FACIAL FIRST IMPRESSIONS

Table 1.

Individual observer factor analyses.

Each row represents the factor analysis solution for an individual observer and shows the percentage of variance explained. The numbers of factors included were determined independently by four criteria used in previous group-based studies (see text for more details). Factors were labelled as trustworthiness (red), youthful/attractiveness (green) or competence/dominance (blue; see text for criteria). Unpredicted factors (orange) are labelled by traits loading most highly. Numbers in brackets indicate the order of factors for each individual participant in terms of the proportion of variance explained (highest = 1).

	Factor numbers				Total variance explained	Factor label (position)			
	Kaisers	Scree	Parallel	MAP		Trust.	Youthful/ Attract.	Competence/ Dominance	Other
1	3	3	3	2	52%	-	youth/attract (2)	dominance (3)	comp/trust (1)
2	4	2,4	3	3	56%	trust (1)	youth (3)	-	att/dom/comp (2)
3	3	2,3	3	3	65%	trust (1)	youth/attract (3)	competence (2)	-
4	3	3	3	3	64%	trust (2)	youth/attract (1)	competence (3)	-
5	3	2-4	3	3	53%	trust (1)	youth/attract (3)	competence (2)	-
6	3	1,4	3	3	62%	trust (1)	youth/attract (3)	comp/dom (2)	-
7	3	2,3	2	2	48%	trust (1)	youth/attract (2)	comp/dom (3)	-
8	3	2,3	2	2	65%	trust (1)	youth/attract (3)	competence (2)	-
9	3	3	3	3	63%	trust (1)	youth/attract (3)	comp/dom (2)	-
10	3	1,3	3	3	54%	trust (1)	youth/attract (3)	dom/comp (2)	-
11	4	4	3	2	69%	trust (1)	youth/attract (2)	competence (3)	dominance (4)
12	3	1,3	2	1	46%	trust (1)	youth/attract (2)	comp/dom (3)	-
13	4	1,2,4	2	1	43%	trust (1)	youth/attract (2)	-	health/intel. (3), att/fem. (4)
14	3	3	3	3	68%	trust (1)	youth/attract (2)	comp/dom (3)	-
15	2	2	2	3	64%	trust (1)	youth/attract (2)	-	-
16	3	1,3, 5	3	2	52%	trust (2)	youth/attract (1)	competence (3)	-
17	3	1,3	3	2	55%	trust (1)	youth/attract (2)	-	health/intel. (3)
18	5	1,3	3	2	48%	trust (1)	youth/attract (2)	dominance (3)	intel. (4), trust (5)
19	3	3	3	3	55%	trust (1)	youth/attract (2)	comp/dom (3)	-
20	3	3,4	3	3	61%	trust (3)	youth/attract (1)	dom/comp (2)	-
21	3	3	3	3	62%	trust (3)	youth (2)	comp/dom(1)	-
22	3	2-5	2	1	33%	trust (1)	youth/attract (3)	comp/dom (2)	-
23	4	2-5	3	2	45%	trust (1)	youth/attract (3)	competence (2)	-
24	3	3	3	3	60%	trust (1)	youth/attract (2)	comp./dom. (3)	-

FACIAL FIRST IMPRESSIONS

Table 2.

Means and SDs of factor loadings across participant models.

Mean and standard deviations of factor loadings across individual factor models. Factors were averaged together if labelled as trustworthiness (23 observers), youthful/attractiveness (24 observers) or competence/dominance (20 observers, see text for criteria). Where participants had multiple factors with the same label, we took the first factor only. Mean factor loadings above .3 are shown in bold.

	Trustworthiness mean (SD) loading	Youth-Attract mean (SD) loading	Comp/Dom mean (SD) loading
Warm.	.80 (.18)	.19 (.12)	.12 (.21)
Approa.	.78 (.13)	.24 (.16)	.15 (.24)
Trust.	.67 (.17)	.21 (.16)	.24 (.23)
Smiling	.75 (.13)	.17 (.11)	.02 (.24)
Attrac.	.36 (.14)	.62 (.22)	.28 (.23)
Health	.34 (.15)	.54 (.22)	.27 (.27)
Age	-.01 (.09)	-.70 (.19)	.26 (.23)
Babyf.	.24 (.21)	.56 (.28)	-.17 (.24)
Dom.	-.24 (.26)	-.13 (.23)	.41 (.26)
Intel.	.32 (.18)	.01 (.21)	.58 (.22)
Masc.	-.29 (.12)	-.53 (.14)	.10 (.24)
Comp.	.28 (.23)	.06 (.21)	.62 (.26)

FACIAL FIRST IMPRESSIONS

Figure captions

Figure 1. A) Matrices depicting agreement between the currently-obtained dimensions (at the individual level) with independently-obtained consensus-based dimensions of trustworthiness, youthful-attractiveness and dominance taken from Sutherland et al. (2013), for each observer separately. Dimensions are sorted in order: T = trustworthiness, YA = youthful attractiveness, D = dominance, C = competence, O = other so that on-diagonal correlations reflect agreement in hypothesised dimensions. Pale green: absolute Pearson's $r = -1$, blue-green: absolute Pearson's $r = 0$, Dark blue: absolute Pearson's $r = 1$. The bottom rightmost matrix depicts a hypothetical ideal observer, who shows complete agreement in their dimensions, relative to independent consensus dimensions (assuming perfect reliability, i.e. $r = 1$ on the diagonal). B) Agreement between individual and independent consensus dimensions (taken from Sutherland et al., 2013) on average across observers ** $p < .001$ * $p < .05$. Error bars represent 95% confidence intervals.

Figure 2. Relative ordering of dimensions across participants. Dimension order is based on variance explained, with lower dimensions explaining more variation. Significance values based on planned X^2 comparisons: ** $p < .001$ * $p < .05$, † $p = .071$.

Figure 3. A) Matrices depicting the strength of agreement between facial impressions of different traits for individual observers (participants 1-24). Traits in each matrix are ordered by trustworthiness, youthful-attractiveness and then competence/dominance factors (from top-bottom and left-right): warmth, approachability, trustworthiness, smiling, attractiveness, health, age, babyfacedness, dominance, masculinity, competence and intelligence. Borders are shown around the regions involving the four traits hypothesised to load particularly on each dimension. B) A matrix showing the strength of agreement in the pattern of these impressions across all observers, with each cell depicting a single pairwise correlation between two different observers (N = 1-24, from top-bottom and left-right; NB: none of the observer pairs showed negative agreement). Palest green: absolute Pearson's $r = 0$, dark blue: $r = 1$. Cells along the diagonal (where the correlation is by definition 1.0) are left blank in all matrices.

FACIAL FIRST IMPRESSIONS

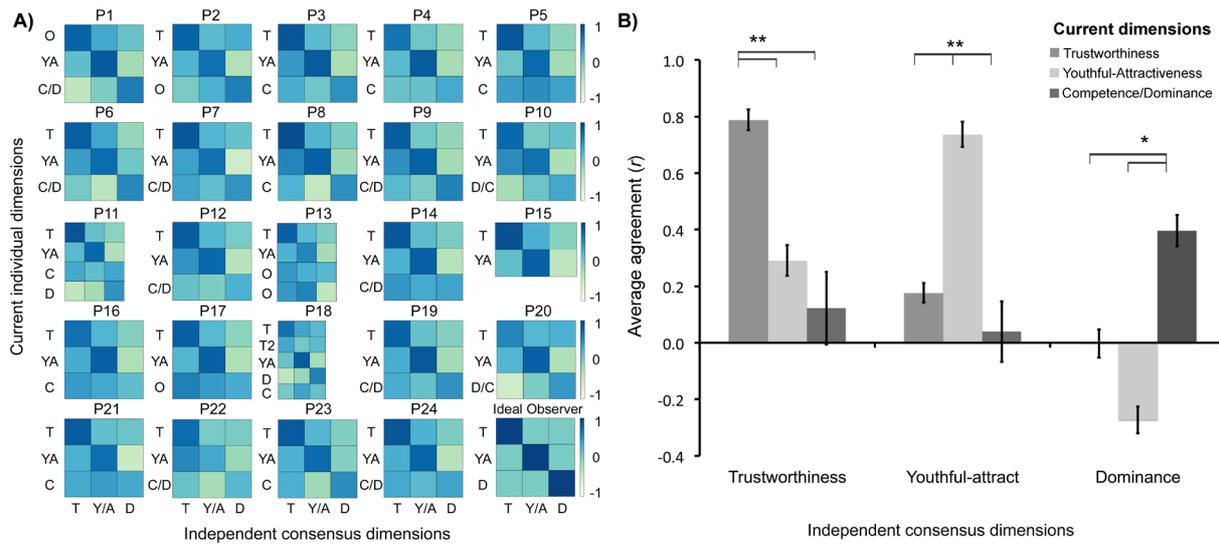


Figure 1. A) Matrices depicting agreement between the currently-obtained dimensions (at the individual level) with independently-obtained consensus-based dimensions of trustworthiness, youthful-attractiveness and dominance taken from Sutherland et al. (2013), for each observer separately. Dimensions are sorted in order: T = trustworthiness, YA = youthful attractiveness, D = dominance, C = competence, O = other so that on-diagonal correlations reflect agreement in hypothesised dimensions. Pale green: absolute Pearson's $r = -1$, blue-green: absolute Pearson's $r = 0$, Dark blue: absolute Pearson's $r = 1$. The bottom rightmost matrix depicts a hypothetical ideal observer, who shows complete agreement in their dimensions, relative to independent consensus dimensions (assuming perfect reliability, i.e. $r = 1$ on the diagonal). B) Agreement between individual and independent consensus dimensions (taken from Sutherland et al., 2013) on average across observers ** $p < .001$ * $p < .05$. Error bars represent 95% confidence intervals.

FACIAL FIRST IMPRESSIONS

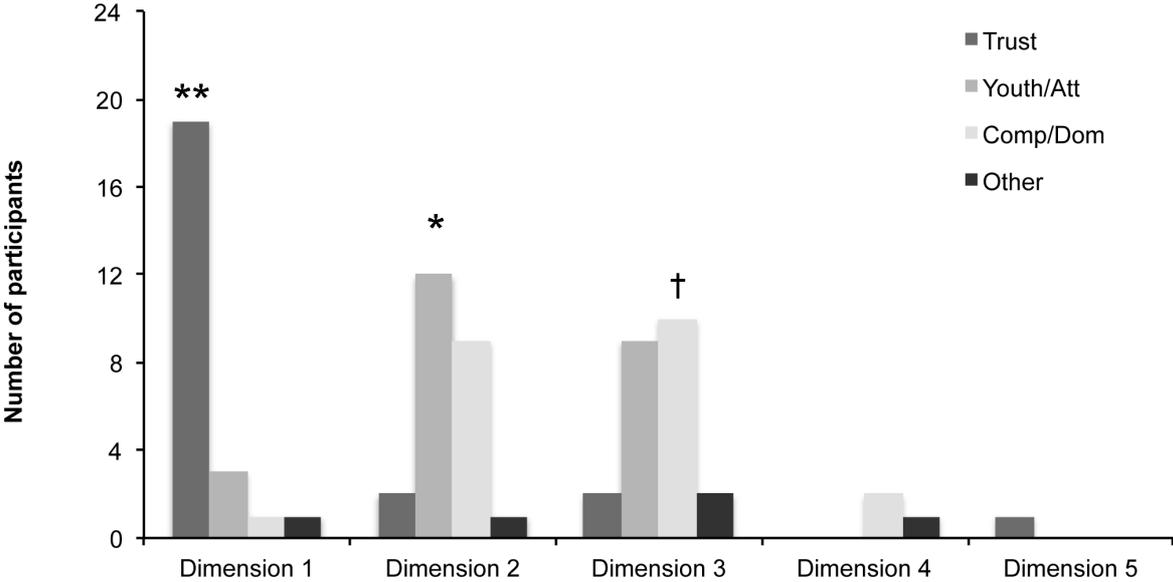


Figure 2. Relative ordering of dimensions across participants. Dimension order is based on variance explained, with lower dimensions explaining more variation. Significance values based on planned χ^2 comparisons: ** $p < .001$ * $p < .05$, † $p = .071$.

FACIAL FIRST IMPRESSIONS

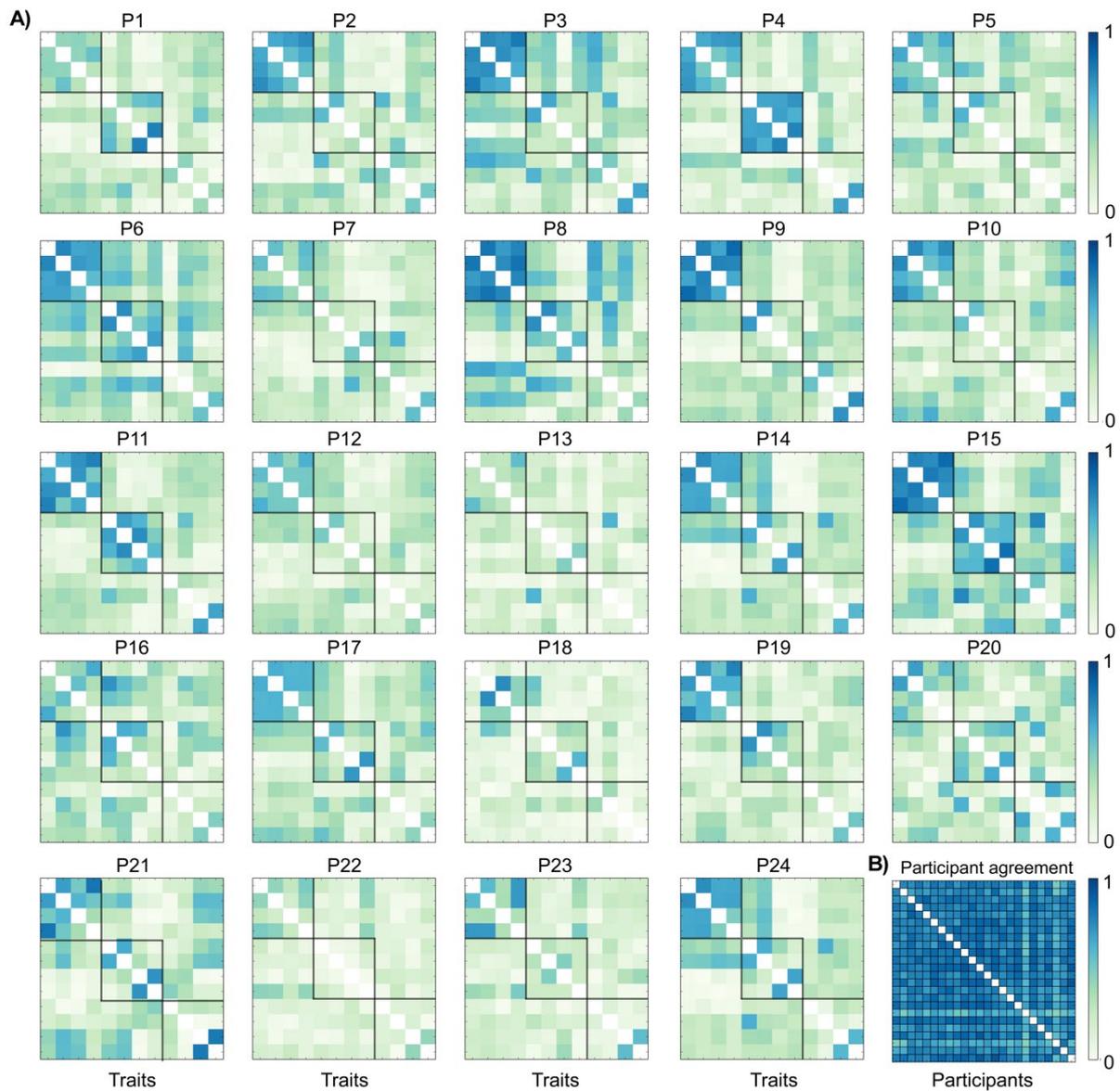


Figure 3. A) Matrices depicting the strength of agreement between facial impressions of different traits for individual observers (participants 1-24). Traits in each matrix are ordered by trustworthiness, youthful-attractiveness and then competence/dominance factors (from top-bottom and left-right): warmth, approachability, trustworthiness, smiling, attractiveness, health, age, babyfacedness, dominance, masculinity, competence and intelligence. Borders are shown around the regions involving the four traits hypothesised to load particularly on each dimension. B) Matrix depicts the strength of agreement in the pattern of these impressions across all observers, with each cell depicting a single pairwise correlation between two different observers (N = 1-24, from top-bottom and left-right; NB: none of the observer pairs showed negative agreement). Palest green: absolute Pearson's $r = 0$, dark blue: $r = 1$. Cells along the diagonal (where the correlation is by definition 1.0) are left blank in all matrices.