Personality in faces: Implicit associations between appearance and personality

Alex L. Jones[1], Jeremy Tree[1], & Robert Ward[2]

[1]Swansea University

[2]Bangor University

Author Note

Corresponding author:
Robert Ward
Wolfson Centre for Clinical and Cognitive Neuroscience
School of Psychology; Bangor University
Brigantia Building
Penrallt Road
Bangor, LL57 2AS
E-mail: r.ward@bangor.ac.uk
The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Abstract

How accurate are the spontaneous trait inferences made to faces? Here we measured implicit associations between facial appearance and personality traits, using faces conveying an objective appearance of Extraversion and Agreeableness. In the standard or "uncrossed" conditions of Experiment 1, we found that descriptions of high and low Agreeableness and Extraversion were spontaneously and accurately associated with their objective trait appearance. In Experiment 2, to test the specificity of this effect, we "crossed" the IATs, pairing faces conveying high and low Extraversion with words describing characteristics of high and low Agreeableness, and the reverse. We found evidence for associations specific to objective appearance of Agreeableness, and a general halo effect relating to Extraversion. We conclude that spontaneous assessment of personality from faces can be accurate, and can be based on trait-specific as well as general visual cues.

Keywords: Personality, Accuracy, Facial Appearance, Implicit Cognition

Personality is spontaneously and accurately detected from the face

Judgements made on the basis of facial appearance can have important outcomes. Facial appearance influences political decisions (Little, Burriss, Jones, & Roberts, 2007), hiring decisions (Luxen & Van De Vijver, 2006), and is even correlated with military career progression (Mazur, Mazur, & Keating, 1984). Judgments of character from facial appearance show a high degree of agreement between observers (Kenny, Albright, Malloy, & Kashy, 1994; Walker & Vetter, 2016), even to the extent that judgements made by children, from around age three, can agree with those made by adults (Cogsdill, Todorov, Spelke, & Banaji, 2014). Given the prevalence, importance, and agreement surrounding these judgements of character, two questions naturally arise. First, are personality inferences accurate when made on the basis of mere facial appearance? Second, if these judgements are accurate, do they happen spontaneously, or only upon explicit instruction? We consider these questions in turn.

Can personality inference based on facial appearance be accurate?

Accuracy of first impressions is often measured by the agreement between the self-report of targets about their personal traits, and the estimates of those traits as made by observers. By this standard definition, many studies have found accurate trait inferences from facial appearance, for a diverse range of traits, including sociosexuality (Boothroyd, Cross, Gray, Coombes, & Gregson-Curtis, 2011; Boothroyd, Jones, Burt, DeBruine, & Perrett, 2008), dominance (Quist, Watkins, Smith, DeBruine, & Jones, 2011), fighting ability (Little, Třebický, Havlíček, Roberts, & Kleisner, 2015), trustworthiness (Tognetti, Berticat, Raymond, & Faurie, 2013), political affiliation (Rule & Ambady, 2010), and both physical (Jones, 2018) and mental health issues (Ward & Scott, 2018), including depression (Scott, Kramer,

Jones, & Ward, 2013) and borderline personality disorder (Daros, Ruocco, & Rule, 2016).

However, concerns have been raised about how to interpret accuracy from facial appearance. In particular, Todorov and Porter (2014) have noted the importance of image selection, and the possibility of confounding cues. Comparing mug shots of felons with publicity photos of Nobel prize winners might produce accurate discrimination, but not necessarily because of facial cues. A related point - and anyone who has seen an unflattering photo of themselves would surely agree - vagaries of lighting, pose, and other factors mean different attributions can be given to different photographs of the same individual (Todorov & Porter, 2014). Again, from the universe of possible images that could be made of a face, we need to ask, how were the photos used as stimuli selected? For example, stimulus images taken from a social networking site might reflect the impression management of the photo targets (Siibak, 2009). Likewise, publicity photos of politicians may be selected to convey messages related to party affiliation (Olivola, Sussman, Tsetsos, Kang, & Todorov, 2012). An additional and related concern is that a number of controllable cues, such as pose, clothing, and hairstyle can be present in facial photos (e.g., Mazur, 2005), and these controllable cues can themselves provide accurate information about an individual (Naumann, Vazire, Rentfrow, & Gosling, 2009). Furthermore, the human face also contains an array of dynamic expression cues relevant for social perception, which are at least under partial control. These sorts of dynamic and controllable cues can be manipulated by the target to produce different impressions on observers, again raising the issue of how to best select images when studying the accuracy of trait inference.

Due to these concerns, many studies investigating trait inference from facial appearance use highly constrained stimuli. In many experiments, judgements of character are made on the basis of static photographs taken by the experimenters, with cues from expression, pose, hairstyle, clothing, and cosmetics eliminated or minimized, typically in the form of "passport"images with neutral expression. However, the use of highly constrained images should not be seen as an attempt to create a "ground truth" of appearance. Consider that having one's photograph taken creates a social context, and as such, there remains scope for some individual variability to emerge. For example, individuals could differ in the posture of their head, mouth, eyes, and other areas, from both voluntary and involuntary responses to the demands of the social situation. Systematic appearance differences arising within the context of highly constrained images therefore might not arise in all other contexts. Therefore, highly constrained image contexts are valuable, not because they represent a ground truth of appearance, but because they minimize controllable cues, and allow for a replicable stimulus creation procedure. It is therefore relevant that even using highly constrained stimuli, accurate inference of personality can be made. A growing body of work has confirmed that even under highly constrained conditions, aspects of the Big Five (i.e., the Five-Factor Model; FFM), including Agreeableness, Extraversion, and Neuroticism (Jones, Kramer, & Ward, 2012; Robin S. S. Kramer & Ward, 2010; Little & Perrett, 2007; Penton-Voak, Pound, Little, & Perrett, 2006) can be reliably detected from facial appearance.

Much of the work assessing trait accuracy in faces has used the trait composite method, which involves averaging the faces of individuals high or low on FFM traits. Composite trait images might be referred to as a genuine or *objective trait appearance*: through averaging, statistical regularities in the appearance of people

sharing a trait would be preserved, while idiosyncrasies minimised or removed. If there were no regularities in appearance, then composites images from one trait extreme would not be reliably distinguishable from composites made from the other extreme. However, personality accuracy has also been shown in individual face images, under both unconstrained (Borkenau, Brecke, Möttig, & Paelecke, 2009) and constrained (Penton-Voak et al., 2006) presentations.

Are accurate personality inferences from facial appearance made spontaneously, or only under explicit instruction?

The studies we have reviewed demonstrating accurate trait inference follow a similar method, presenting observers with faces and explictly asking the observers to make a judgement about the faces (e.g., Penton-Voak et al., 2006). Our next question is therefore whether accurate inferences arise spontaneously, or are limited to cases of explicit instruction. To summarise the position we develop below, we believe the current literature does not answer this question. With limited exception, previous studies looking at spontaneous trait inference from facial appearance do not consider accuracy, and studies looking at accuracy do not consider spontaneous trait inference.

There is a long history in social psychology research documenting spontaneous trait inference, in which observers make social judgements about others even when not explicitly instructed to do so (J. S. Uleman, 1987). Although this history of research is sometimes said to demonstrate that trait inference is "automatic", it is important to be clear about the meaning of this term. For example, some previous studies have claimed to demonstrate that social attributions to faces are made in an "automatic" way, because inference occurs so rapidly (Ballew & Todorov, 2007; Cogsdill et al., 2014; Rule, Ambady, & Hallett, 2009). Indeed, multiple demonstrations show that attributions to faces can be made from brief exposures of

approximately 50-100 ms (Borkenau et al., 2009; Olivola & Todorov, 2010; Olson & Marshuetz, 2005; Rule & Ambady, 2008; Rule et al., 2009; Willis & Todorov, 2006) and that unspeeded, deliberate contemplation of attributions produces similar results as instructions to make inferences on the basis of first impressions or gut feeling (Ballew & Todorov, 2007; Rule et al., 2009).

However, "automatic" can have many meanings with regard to cognitive processes (e.g., Bargh, 1989; Kahneman & Tresiman, 1984), with some of the key distinctions being made between rapid and slow; capacity-limited and capacity-free; and spontaneous and deliberative processes. For example, a process might be rapid, but made only under deliberate control. In the studies above, showing rapid "automatic" inference, the task for participants was to explicitly report attributes of the faces presented. These studies therefore do not indicate whether accurate social inference from faces can be "automatic" in the sense of being made spontaneously and without specific task instruction.

Of course, there are many studies which do consider spontaneous trait inference from facial appearance, and its pervasive influence. We mentioned some at the start of our introduction (Little & Perrett, 2007; Luxen & Van De Vijver, 2006), but there are many others (for review of spontaneous inference, see Uleman, Adil Saribay, & Gonzalez, 2008). However, in general, previous studies on spontaneous inference to appearance do not, or can not, consider the accuracy of inferences made. We illustrate this point with three typical examples. Todorov et al. (2005) found that voters chose politicians that other raters found to be competent-looking. This appears to be a spontaneous use of appearance information by voters, as they are not instructed to rate politicians on competent appearance. However, lacking any measure of target competency, we cannot infer that either the voting behaviour or the ratings of

competence reflected accurate judgements of competence. Similarly, consider Zebrowitz and McDonald's (1991) famous observations on facial appearance and litigation outcomes. One of several effects they observed was that attractive plaintiffs were more likely to prevail. This effect was spontaneous in the clear sense that judges were not asked to evaluate attractiveness. But we have no way of knowing whether the legal decisions were objectively correct, and in fact, Zebrowitz and McDonald suggest these spontanous inferences were certainly extralegal, and probably inaccurate and biased. Finally, attributions consistent with an attractiveness halo appear to be drawn spontaneously, in that attractive faces were associated with all manner of positive words (van Leeuwen & Macrae, 2004). However, spontaneous attributions of generally positive characteristics to attractive facial images are again unlikely to reflect accurate judgement, but a biased attractiveness halo (Eagly, Ashmore, Makhijani, & Longo, 1991).

To our knowledge, there are only limited demonstrations for the accuracy of spontaneous inferences to facial appearance, particularly when appearance is highly constrained. The demonstrations we are aware of apply to the trait of trustworthiness. Accurate judgements of trustworthiness in economic games, defined as in-game responses benefitting the observer, can be made from "thin slices" and brief interactions (e.g., Sparks, Burleigh, & Barclay, 2016). That is, without instruction to do so, game players have been found to use information from the appearance of their partners to improve the predictions of their partners' actions. This spontaneous yet accurate use of appearance information has also been applied to highly constrained facial appearance (see Bonnefon, Hopfensitz, & Neys, 2017, for a review). Stirrat and Perrett (2010) found that facial width in men was a valid cue for behaviour in a trust game, and that men with relatively wide faces were spontaneously evaluated as less

trustworthy in this game. In this case, the perceptions of trustworthiness for individual targets was not compared to the behaviour of targets, as the two correlations were based on separate groups of men. Bonnefon, Hopfensitz, and De Neys (2015) found that while full-face color pictures of partners did not promote significant accuracy, black-and-white images, cropped to an area around the internal facial features, did. The evidence for spontaneous yet accurate inference of trustworthiness from facial appearance is therefore suggestive, but limited. Furthermore, we are not aware of any demonstrations for the spontaneous yet accurate inference of personality from constrained facial appearance. Our original question for this section -- are accurate trait inferences from facial appearance made spontaneously – is therefore underexplored.

The Current Study

Here we investigate spontaneous associations to objective trait appearances of personality. In particular, we measure associations made to highly constrained composite images, created from women scoring high and low on self-reported personality measures for Extraversion and Agreeablenes. We use a novel version of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) we will refer to as the Personality in Faces IAT (PIF-IAT). The methodology of the IAT is well known, measuring spontaneous and readily available associations between objects and concepts. Associations that are available without explicit task instruction, will increase performance in one of the IATs critical blocks where a category object and a concept require a response mapped to the same key, compared to when they are reversed.

The literature on the IAT as a measure of implicit cognition is extensive and varied. On one hand, the IAT is argued to measure unconscious yet automatic

associations relevant to social cognition (Greenwald et al., 1998), with the great advantage that associations are not the result of direct self-report. On the other hand, the IAT has received extensive criticism often levelled at its construct validity (Blanton, Jaccard, Christie, & Gonzales, 2007). An important question is to what extent the IAT measures associations genuninely outside conscious awareness (Olson, Fazio, & Hermann, 2007). Although people are often surprised by their IAT results (Mitchell, Nosek, & Banaji, 2003), it has also been found that participants can sometimes accurately predict their IAT score towards different social groups, making it difficult to verify whether the IAT is measuring unconscious as opposed to consciously available associations (Hahn, Judd, Hirsh, & Blair, 2014). It is also unclear whether and to what extent an individual's IAT score reflects their personal, as compared to normative, cultural, or other extrapersonal associations (Arkes & Tetlock, 2004; Gawronski, LeBel, & Peters, 2007), although this does not necessarily undercut many experimental and applied uses of the IAT (Uhlmann, Poehlman, & Nosek, 2012). Research and debate therefore continues on many aspects of the IAT, including whether it measures truly unconscious processes.

However, for our adaptation of the PIF-IAT, we were motivated to draw on two other methodologically less controversial aspects of the IAT, which are nonetheless crucially important for our work. One is the structure of the IAT task: participants are never explicitly asked or instructed to rate the concept stimuli for the presence of the attribute. For example, in a racial IAT, participants are never explicitly asked how different positive and negative attributes might be related to the concepts of black and white people (Smith-McLallen, Johnson, Dovidio, & Pearson, 2006). Any association measured is therefore due to a spontaneous association of stimuli and responses, whether through a mechanism like priming or strategic

recoding (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009) A second aspect of the IAT which is important for us lies with core principles of associative learning – people find it easier to give the same response to stimuli that are strongly associated, compared to those that are weakly associated (Nosek, Greenwald, & Banaji, 2005). Or equally, people find it more difficult to give different responses to strongly associated, compared to weakly associated, stimuli. Conceptual analyses have suggested this linkage between associated stimuli and responses could arise due to multiple mechanisms, including response priming between associated stimuli, and differential task switching costs relating to the overlap of attribute and concept features (De Houwer et al., 2009). We argue that these aspects of the IAT make it an appropriate and potentially effective way to measure spontaneous associations between facial appearance and personality traits.

Traditional IAT approaches using faces as a category have a simple superordinate label that facilitates categorization, such as race (black or white; McConnell & Leibold, 2001) or age (old or young; Nosek et al., 2007). However, the face stimuli we use here are not easily described without an explicit mention of their personality types. We instead cast the faces category as an identity recognition task. We name one composite “Jane” and the other “Mary”. Participants respond to a simple identity-sorting task (*Jane, Mary*) as face stimuli appear on screen, and no mention needs to be made of any relationship between the faces and their corresponding personality traits. For testing associations of faces to personality concepts, we used words describing high or low levels of a persoanlity trait (e.g., (*kind*, *sympathetic, helpful, warm; versus cold, unsympathetic, harsh, unkind*). Because the composite faces we use represent objective trait appearances, any

association between a composite and its trait-congruent personality words represents a spontaneous yet accurate association of appearance and personality.

Experiment 1: Spontaneous associations to objective trait appearance

In our first experiment, we tested whether faces signaling Extraversion and Agreeableness were spontaneously associated with those traits. We focused on two traits from the Five Factor Model (FFM; McCrae & Costa, 1989), Agreeableness and Extraversion. These two traits have been repeatedly shown to be cued from faces in studies using explicit report methods (Jones et al., 2012; Robin S. S. Kramer & Ward, 2010; Little & Perrett, 2007; Penton-Voak et al., 2006). The associations measured are therefore spontaneous in the direct sense that they are not instructed. Accuracy will be determined by whether there is a bias towards congruent over incongruent face-word associations.

Method

Participants

We utilized a between groups design, collecting separate samples for each of the IATs, with a total sample of 227. We used G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) to conduct a sensitivity analysis for our studies, and our ability to detect an association if present in each of the IAT tests, as indicated by a bias signficantly greater than zero in a one-sample *t*-test. With a minimum of $n = 94$ in each test, $\alpha = 0.05$ and $\beta = 0.80$, we expected to detect effect sizes of $d > .3$, conventionally a medium effect. The samples are described below in full. Details of participants that were excluded due to an excessive number of very fast responses after application of the revised scoring algorithm (participants removed if 10% or more of trials were less than 300 ms, as outlined by Greenwald et al., 2003) are also included. Ethics approval for the study was obtained from Swansea University.

Extraversion IAT. For this PIF-IAT, there were 94 participants (age *M* = 26.38, *SD* = 10.17, 57 females). Participants were recruited through Prolific.ac, and were compensated £1.25 for their participation. Three male participants were removed after scoring the data, for a final sample of *n* = 91.

Agreeableness IAT. One hundred and twenty three participants (age unknown for two participants, age *M* = 27.45, *SD* = 11.01; 64 females) completed this PIF-IAT. Participants were recruited for course credit or through social media. One male participant was removed after scoring the data, for a final sample of *n* = 122.

Stimuli

Four facial composites used in previous research (Robin S. S. Kramer & Ward, 2010) served as stimuli. The composites were generated from a sample of 64 Caucasian females (age *M* = 21.03, *SD* = 1.94), who posed for a neutral facial photograph before completing the Mini-IPIP (Donnellan, Oswald, Baird, & Lucas, 2006), a 20-item measure of the Big Five personality factors. The 15 highest and lowest scorers for Extraversion and Agreeableness were identified, and their facial photographs were averaged using Abrasoft Fantaface Mixer. The standardised scores for each composite image on Extraversion and Agreeableness are provided in Table 1.

Table 1. Mean Agreeableness and Extraversion scores by stimulus

Composite image	Extraversion	Agreeableness
Low Extraversion	-1.39	-0.17
High Extraversion	1.17	0.07
Low Agreeableness	0.13	-1.38
High Agreeableness	0.41	1.14

Note. Standardised mean trait scores ($M = 0$, $SD = 1$) for the fifteen women within each composite image. Scores were first standardised for the sixty-five women in the photographic database (e.g., the women in the low Agreeableness composite had an average Agreeableness score 1.38 SDs below the group mean).

Across all PIF-IAT studies, high Agreeableness and Extraversion faces were named ‘Jane’ and their low counterparts were named ‘Mary’. Composite faces are shown in Figure 1.

Figure 1. The composite faces used in the study. Low-level trait composites appear in the left column (named ‘Mary’ in the IAT), and high-level trait composites appear on the right (named ‘Jane’ in the IAT). Top row: Agreeableness, bottom row: Extraversion.

Procedure

Participants completed only one of the PIF-IAT s, following the block structure described in Table 1. Before completing the task, participants were presented with a description of the categories in the experiment, and were shown the facial composites (labeled *Jane* and *Mary*) and the high and low trait words and their categories. Participants were told the experiment was a simple sorting task, matching faces to the correct identity with a keypress, and words that matched personality descriptions. They were given no further information regarding the nature of the composites. Before the trial presentations, participants were shown the images of Jane and Mary at a larger resolution (445 x 485) for a fixed duration of two minutes. Participants were instructed to learn the identities of Jane and Mary in preparation for the subsequent sorting task, and were unable to continue until the time had elapsed. A practice block immediately followed this familiarization task, wherein participants completed four runs of ten trials categorizing Jane and Mary. In the first half of these trials, Jane appeared on the left and Mary on the right, before the order was switched. The purpose of this practice block was to further familiarize participants with Jane and Mary under the usual response conditions of the IAT.

Following standard IAT procedures, participants completed the seven blocks, with the orderings of "congruent" and "incongruent" blocks reversed for 48.18% of participants across both IATs. For our tasks, a congruent block was defined as responses to Jane (the composite made of high scorers on that trait) being on the same key as words describing high levels of that trait. Category labels (e.g., Jane, high Extraversion) appeared on the top left and right of the screen. Participants responded by pressing the "E" key for a left response, and the "I" key for a right response. If an error was made, a red cross appeared underneath the current stimulus, and participants

had to correct their response. In test blocks, each individual word appeared twice, and images of Jane and Mary eight times each.

Both tests were conducted online, and participants completed the task over the Internet. The IAT has been used extensively as a web-based experiment successfully (e.g., Project Implicit; Greenwald et al., 2003), and so we considered a non-laboratory sample suitable for this study.

Table 2. An outline of the uncrossed Extraversion IAT

Block	Trials	Function	Left key response	Right key response
1	16	Practice	Jane (high Extraversion composite)	Mary (low Extraversion composite)
2	16	Practice	High Extraversion words	Low Extraversion words
3	32	Test	Jane & high Extraversion words	Mary & low Extraversion words
4	32	Test	Jane & high Extraversion words	Mary & low Extraversion words
5	16	Practice	Mary (low Extraversion composite)	Jane (high Extraversion composite)
6	32	Test	Jane & low Extraversion words	Mary & high Extraversion words
7	32	Test	Jane & low Extraversion words	Mary & high Extraversion words

Note. As is standard for IAT procedures, blocks 1, 3, and 4 are switched respectively with blocks 5, 6, and 7, to vary the order in which congruent (shown) and incongruent trials appear. In the crossed version of this IAT, high and low Agreeableness words appear alongside Extravert composites. The reverse is true for Agreeableness composites.

Results

Reaction time data were converted to IAT-D scores, which are a form of effect size measure, comparing the latencies in congruent to incongruent conditions. A positive D score in our case reflects a bias to make the congruent association; for example, high extraversion composites with high extraversion words. The D-scores were calculated according to the revised scoring algorithm described by Greenwald et

al (2003). As a correct response was required after a incorrect trial, we added the time taken to provide a correct response to the initial reaction time as an error penalty.

We conducted an initial one-sample *t*-test against zero to test for the significance of any association between the faces and trait words. We found a significant D score for Extraversion, D = 0.29, 95% CI [0.21, 0.37]), $t(90) = 7.28$, $p < .001$, $d = 0.76$; and for Agreeableness, D = 0.30, [0.23, 0.38]), $t(121) = 8.14$, $p < .001$, $d = 0.74$. Facial composites of personality traits were therefore spontaneously and accurately associated with corresponding trait words.

Experiment 1b: Eliminating naming confounds

Our initial results indicate that associations to faces conveying actual personality information are accurate and can occur spontaneously. However, a possible confound is our use of fixed category labels for each image. For example, the low Agreeableness composite was always called ‘Mary’. It is possible that these names may drive the accurate and spontaneous associations shown. Although we have no reason to expect that Mary and Jane might differ significantly in the associations they drive, and while these names are relatively high-frequency and unremarkable, it is nevertheless certainly conceivable that the names, rather than the faces, might be driving the associations: for example, ‘Jane’ might be perceived as a more friendly or outgoing name than ‘Mary’. We therefore carried out a conceptual replication of Experiment 1 with the name labels swapped. That is, high level trait composites were now named ‘Mary’, and low level composites were named ‘Jane’. If the names are indeed driving spontaneous trait attributions, the strength and direction of biases should now change.

Indeed, although there is mixed evidence about the degree to which names are linked to appearance (compare Kramer & Jones, 2015; Zwebner, Sellier, Rosenfeld,

Goldenberg, & Mayo, 2017). observers can associate different kinds of names with different personality dimensions (Kramer & Jones, 2015; Zwebner, Sellier, Rosenfeld, Goldenberg, & Mayo, 2017), and pairing faces with more desirable names increases the attractiveness of the face (Garwood, Cox, Kaplan, Wasserman, & Sulzer, 1980). Moreover, names that were more popular in the past (such as those used here) are generally assigned lower ratings on important social traits such as competence (Young, Kennedy, Newhouse, Browne, & Thiessen, 1993).

Method

Procedure and stimuli were identical to Experiment 1, except that the labels associated with each image were now swapped.

Participants. One hundred and sixty-eight additional participants were recruited through Prolific.ac, and were compensated with £1.25. For the name reversal Extraversion PIF-IAT, there were 86 participants (age unknown for ten participants, age $M = 35.58$, $SD = 12.58$; 56 females). Two males were removed after scoring, for a final sample of $n = 84$. An additional 82 participants completed the Agreeableness PIF-IAT (age unknown for nine participants, age $M = 37.82$, $SD = 11.98$; 39 females), with five males removed after scoring, for a final sample of $n = 77$. With a minimum of $n = 77$ in each test, $\alpha = 0.05$ and $\beta = 0.80$, we expected to detect a bias significantly greater than zero, assuming an effect size of $d > 0.3$, as before.

Results

A one-sample *t*-test against zero revealed significant biases for both the sets of stimuli: Extraversion, D = 0.29, [0.20, 0.38], $t(83) = 6.16$, $p < .001$, $d = 0.68$, and Agreeableness, D = 0.28, [0.18, 0.38], $t(76) = 5.53$, $p < .001$, $d = 0.63$, successfully replicating the initial study. Importantly, there were no significant differences

between D scores from Experiment 1a and those collected here, for both Extraversion, $t(178) = 0.13$, $p = .895$, $d = 0.02$, as well as Agreeableness, $t(203) = 0.94$, $p = .349$, $d = 0.13$, ruling out the possibility that the name labels had any particular influence on the accurate and spontaneous attributions already observed.

Experiment 2: Specificity of implicit associations

The results of Experiment 1 suggest that associations to "objective trait appearances", depicting Extraversion and Agreeableness, are accurate and occur spontaneously. However, these associations might have been generated in two ways. First, spontaneous attributions could be based on specific personality information contained in the composite images. For example, the high Agreeableness composite might contain visual cues specific to traits like warmth and empathy, and the low Extraversion composite, specific cues to traits like a reserved nature. Alternatively, spontaneous attributions could be based on a general attractiveness or other halo of social desirability (Dion, Berscheid, & Walster, 1972). For example, by this account, the high Extraversion composite would be attributed an outgoing nature, but also other socially desirable characteristics not directly related to Extraversion, such as warmth and empathy. Of course, these possibilities are not exclusive, and objective trait appearances could contain both general cues to social desirability, and specific cues to the corresponding trait. Furthermore, there is no reason to expect a priori that all personality traits should reveal the same cue structure: some traits might be revealed by specific cues, and others by general ones.

In our second experiment, we assessed to what extent accuracy of spontaneous attributions related to general and to specific cues. This time, rather than pair Extraversion composites with Extraversion words, and Agreeableness composites with Agreeableness words, as in Experiment 1, we crossed the mapping. We paired

Extraversion composites with trait words related to Agreeableness, and Agreeableness composites with trait words related to Extraversion. If accuracy in Experiment 1 were entirely due to *general* cues for social desirability, then the crossed mappings should have little effect on the total bias. That is, significant bias in this "crossed" version of of the PIF-IAT would indicate the trait composites contain general cues to social desirability which transfer to other traits. Alternativey, if accuracy for a trait were due entirely to *specific* personality cues within the composites, then when the composite-word mappings are crossed, no association bias should be found. That is, in this case bias is the result of specific personality cues in the composites which do not transfer to other traits. Finally, if we find both a significant bias in the crossed version, but which is nevertheless reduced relative to the uncrossed version, the implication is that the composite contains both general and personality-specific cues.

Method

Participants. One-hundred and sixty-eight (168) participants were recruited for course credit and through social media. For the crossed Extraversion PIF-IAT (that is, faces signalling Extraversion paired with Agreeableness words), there were 78 participants (age unknown for three participants, age $M = 27.50$, $SD = 12.11$; 46 females). One female participant was removed after scoring the data, for a final sample of $n = 77$. For the crossed Agreeableness PIF-IAT (faces signalling Agreeableness with Extraversion words), there were 90 participants (age unknown for one participant, age $M = 25.02$, $SD = 9.25$; 56 females), with one female participant removed after data scoring, for a final sample of $n = 89$. Power was comparable to Experiment 1a: with a minimum of $n = 77$ in each test, $\alpha = 0.05$ and $\beta = 0.80$, we expected to detect a bias significantly greater than 0, assuming an effect size of $d > 0.3$.

Stimuli and Procedure. All other aspects of the method were the same as in Experiment 1, except that the words corresponding to Extraversion were presented with the Agreeableness composites, and the words corresponding to Agreeableness were presented with the Extraversion composites.

Results

Scoring was identical to Experiment 1. We found a significant positive D score for Extraversion, D = 0.24, 95% CI [0.13, 0.35]), $t(76) = 4.41$ $p < .001$, $d = 0.50$; but not for Agreeableness, D = -0.08, [-0.18, 0.01]), $t(88) = 1.79$, $p = .076$, $d = -0.19$. These results on their own imply that associations to the Extraversion composites reflect a general halo effect, while associations to the Agreeablness composites are consistent with specific visual cues to Agreeableness.

To verify any differences in bias between the "uncrossed" Experiment 1, and the "crossed" version in Experiment 2, we submitted the D scores from both Experiments to a 2 (Trait of Face: Extraversion or Agreeableness) x 2 (Crossover: Uncrossed Experiment 1 or Crossed Experiment 2) mixed ANOVA, using Type III sums of squares to mitigate the unbalanced cell sizes. There was a main effect of Trait of Face, such that Extraversion faces ($M = 0.26$, [0.20, 0.33]) produced higher D scores than Agreeableness faces ($M = 0.11$, [0.05, 0.17]), $F(1, 375) = 12.18$, $p < .001$, $\eta_p^2 = .03$. There was also a main effect of Crossover, such that D scores were higher in the Uncrossed condition ($M = 0.30$, [0.24, 0.35]) than when facial signals did not match the associated personality descriptions in the Crossed condition ($M = 0.08$, [0.01, 0.14]), $F(1, 375) = 23.99$, $p < .001$, $\eta_p^2 = .06$. Finally, both of these main effects should be interpreted in the context of the significant interaction of Trait and Crossover (as shown in Figure 2), $F(1, 375) = 14.73$, $p < .001$, $\eta_p^2 = .04$. To further explore this interaction, we ran contrasts between Crossover conditions for

Extraversion and for Agreeableness. Results for Extraversion were unaffected by Crossover, M_{Diff} = 0.05, [-.08, .17], p = .476, d = 0.11, while there was a difference for Agreeableness, M_{Diff} = 0.39, [.27, .50], p < .001, d = 0.91.

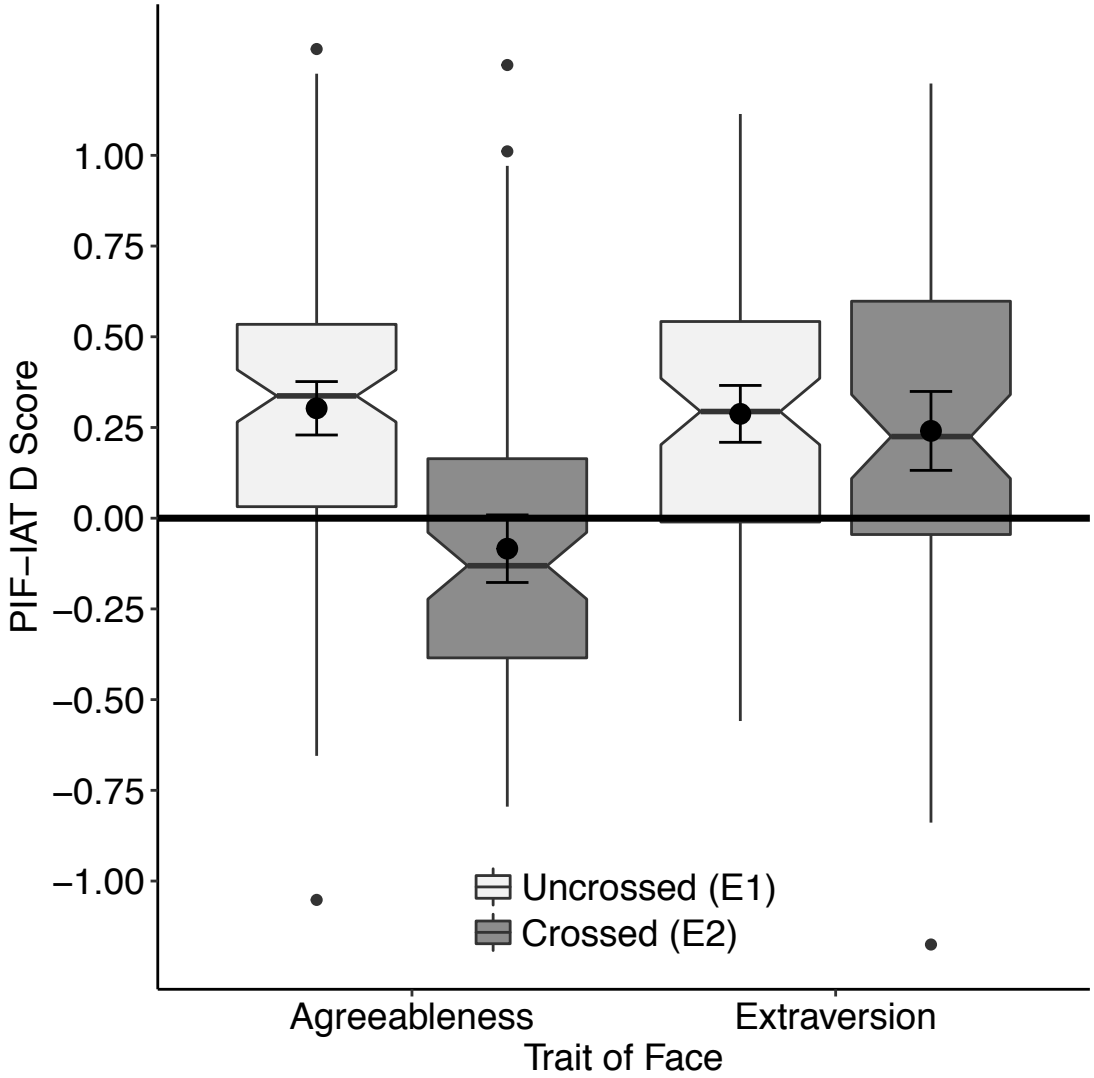

Figure 2. Data summaries across all four PIF-IATs. Black circles represent the average D score, error bars represent 95% CI. Error bars not crossing the zero-line represent a significant bias. Trait indicates the FFM trait conveyed in the faces, either Extraversion or Agreeableness. In uncrossed conditions, D reflects the association of the faces to words of the same trait; in crossed conditions, D reflects associations between faces communicating one trait and words describing the other.

Figure 3 illustrates estimates of the general and specific signals within the Extraversion and Agreeableness composites. The signal for general social desirability is estimated simply as the bias shown in the Crossed conditions, that is, the association between appearance and the social desirability of a different trait (e.g., Agreeable appearance and Extraversion trait adjectives). The specific signal is estimated as the difference in bias between the Uncrossed and Crossed conditions, that is, the strength of assocation of faces with their corresponding trait, which is not explained by a general bias.

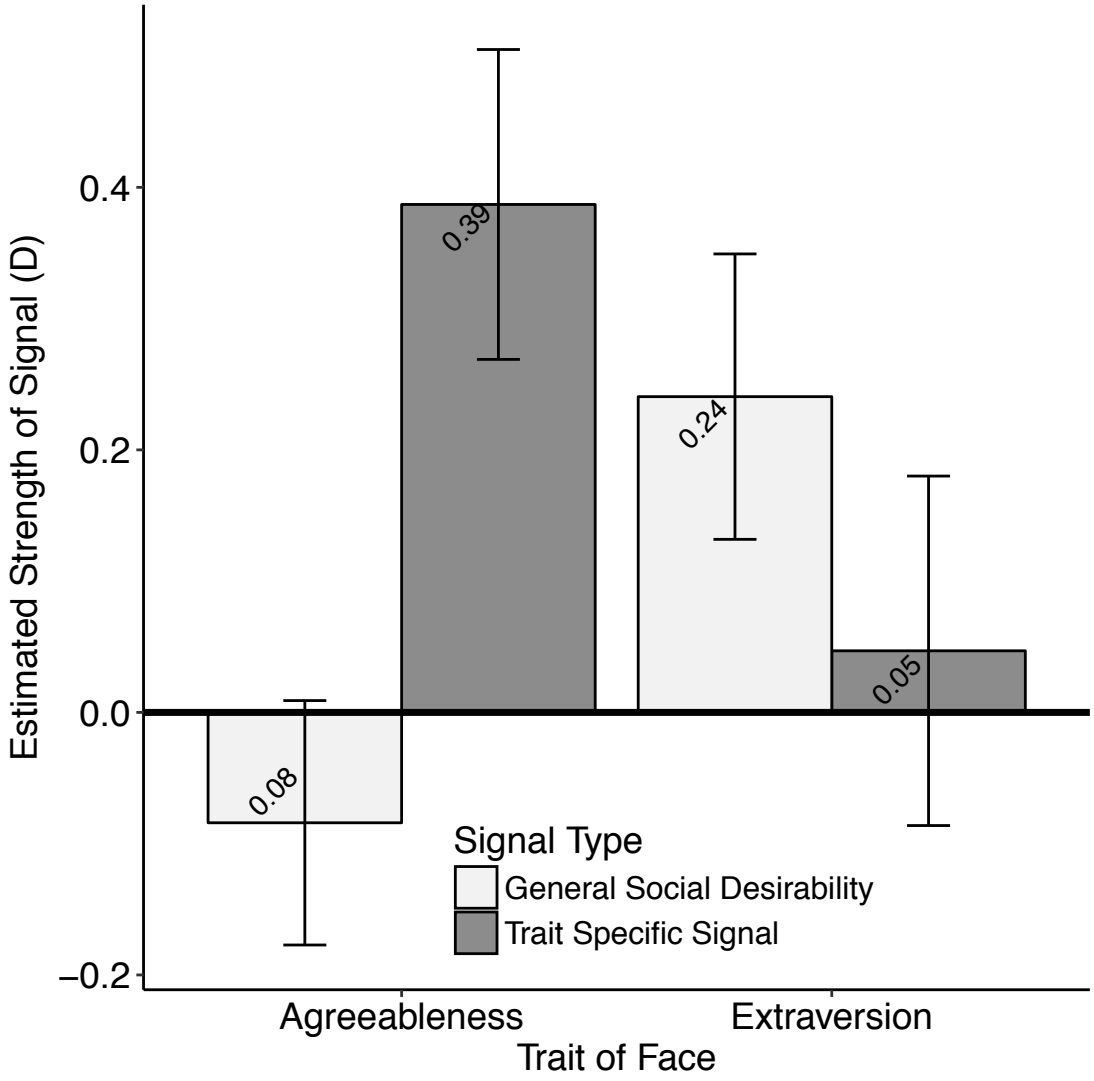

Figure 3. Estimated strength of general and specific signals. General social desirability estimates are the bias the objective trait appearance received under the 'crossed' condition, while trait specific

signals are calculated as the difference between uncrossed and crossed bias scores for each trait appearance.

Our findings suggest that accuracy for the Extraversion and Agreeableness composites are based on very different forms of information. Results from Extraversion can be explained by a general halo effect relating to social desirability of either trait tested. However, visual signals of Agreeableness were specifically associated with the trait of Agreeableness.

General Discussion

We investigated spontaneous associations to objective trait appearances of Extraversion and Agreeableness using a variant of the IAT, the Personality in Faces IAT or PIF-IAT. In the standard, "uncrossed", conditions of Experiment 1, participants correctly and implicitly associated facial composites of women scoring high or low on these traits with corresponding trait adjectives. For example, composites of women scoring high relative to low on trait Agreeableness were more associated with high-agreeable attributes like "friendly" and "warm". After ruling out the possibility that the name labels used drove the effect, the nature of these associations was clarified in the "crossed" conditions of Experiment 2, in which the composites were paired with the adjectives unrelated to their trait (Extraversion composites with Agreeableness words, and vice versa). Here we found evidence of a general halo effect for Extraversion, and a more specific trait association relating to Agreeableness. That is, while Extraversion composites were associated with words describing Agreeableness, Agreeableness composites showed no bias relating to Extraversion. Our interpretation of these findings is that while spontaneous and accurate inference relating to Agreeableness reflects genuine cues to Agreeableness within the composites, the apparent accuracy found in Experiment 1 for Extraversion composites may simply reflect a general association of positive traits to the high

Extraversion composite and negative traits to the low. Taken together, our results indicate that visual cues correlated with personality can be associated with both specific trait perceptions and general social desirability, and that these associations can occur at an implicit, spontaneous level of cognition.

General and trait-specific effects. The halo effect we observed for Extraversion composites is consistent with previous work suggesting that extraverts may be more attractive than introverts (Kramer & Ward, 2010; Pound, Penton-Voak, & Brown, 2007). It is less clear whether facial attractiveness varies greatly as a function of Agreeableness. Attractiveness related to high Agreeableness has been found in spontaneous photos of the head and upper body, although this was tied to controllable cues relating to grooming, not available in the constrained facial images used here (Meier, Robinson, Carter, & Hinsz, 2010). However, most relevant would be previous work with these stimuli by Kramer and Ward (2010), who found a difference in attractiveness between the high and low Extraversion composites, and a smaller but significant difference between the Agreeableness composites. Therefore, a significant difference in rated attractiveness is not sufficient to produce an spontaneous association to all positive traits. This might mean simply that the attractiveness difference for Agreeableness, while significant, was not large enough to drive a general bias to associate the faces with all sorts of socially desirable traits (that is, the halo was not "big enough"). Another speculative hypothesis would be that ratings of facial attractiveness could reflect two factors. First, physically attractive features of the face (e.g., evidence of femininity in women), would drive halo effects and general positive assocations. Second, attractiveness ratings might be influenced in a more specific manner by the social attributions made to that face. For example, a face high in agreeableness might be rated as attractive because it is sending a

desirable social signal. We speculate that these specific social attributions might not as readily generalise to other traits. This speculation simply underscores the point that further investigations are needed to better understand the relationship between attractiveness and social attributions, and in what ways this goes beyond a simple halo. For example, attractive targets may motivate observers to take on a more thorough analysis of the target's social attributes, which may affect impression accuracy (Biesanz, 2010), and impression accuracy may also feedback to affect the observer's ratings of attractiveness (Lorenzo, Biesanz, & Human, 2010).

We can rule out one possibility for the halo effect seen with the Extraversion composites: namely, a confound between Extraversion and Agreeableness scores in the Extraversion composites. The high and low Extraversion composites consisted of women with almost identical mean Agreeableness scores (Table 1).

Potential importance of spontaneous associations. We have already considered some of the debate around the IAT, including whether it is a measure of unconscious processing. We make no claims that the associations we have observed are unconscious, or that they demonstrate a dissociation from explicit and accurate face evaluations; rather we argue our findings indicate these accurate associations may emerge consistently from observers without explicit instruction. Spontaneous inference is an interesting and potentially important phenomenon, which lets us better understand the nature of nonverbal human communication. Claims that targets are "involuntarily broadcasting" a message about their personality to observers (e.g., Scott et al., 2013) really only apply if accurate personality inference occurs spontaneously. There would still be a more limited set of circumstances in which observers might explicitly ask themselves, "Does this person look reserved? Does this person look warm and agreeable?". But the fact that accurate inference seems to be

occurring spontaneously increases the importance of this nonverbal communication channel, and consistent with interpersonal models of trait formation (Joiner & Coyne, 1999), and the possibility that observers may respond negatively to targets on the basis of their inferred personality.

Potential importance of accurate associations. Our results argue that observers can spontaneously make accurate inferences from facial appearance, most clearly demonstrated by the trait-specific signal of Agreeableness. However, the generalised halo we observed for Extraversion composites simultaneously demonstrates how spontaneous associations can be misleading. It is difficult to reconcile the idea of the human brain as a highly functional, well-tuned processor for social information (Alexander, 1990; Little, 2017; Trivers, 2000) with the idea that inaccurate and possibly misleading associations are being routinely drawn from facial appearances (Olivola & Todorov, 2010; Todorov & Porter, 2014). We suggest that associations from appearance are spontaneously drawn, not because observers are always correct, but because they are being rewarded sufficiently often to keep drawing these surface impressions. That is, accuracy from shallow inference does not necessarily need to be high, but better than chance under some circumstances. An evolutionary perspective on communication also reminds us to consider the perspective of the signal sender, in this case the person whose face is being "read". Unless there are benefits, on average, to both the signal sender and receiver, the signal system would not be expected to be maintained over evolutionary time (Maynard Smith & Harper, 2003). For example, senders may benefit by embedding false or manipulative cues within a generally reliable communication channel (Krebs & Dawkins, 1984). From this adaptive perspective, we might therefore expect all sorts of communication channels to demonstrate a mixture of valid and invalid messages,

including communication of personality from facial appearance (e.g., Little, 2017). An important and unresolved question is how to understand the possible adaptive benefits for signalling undesirable social traits (or alternatively, of failing to signal desirable social traits). One possibility is that there may be some social benefit to simply being predictable (Biesanz, 2010). For example, if a target is low in Agreeableness, they might be better off "admitting" that through their appearance, than risking punishment for advertising a trait they do not have (e.g., the increased punishment given to attractive fraudsters, Sigall & Ostrove, 1975).

Accurate trait inference and appearance. Finally, we consider an overarching question that lies behind both this study and related research. Is there variation in facial appearance that is correlated with variation in personality? The fact that observers can accurately discriminate trait levels appears to show this, but here we address two points raised by Todorov and Porter (2014), relating to the accuracy of social impressions from appearance. The first is that accuracy must be considered within the context of image selection, and whether there are biases in this process. One can ask whether the categorisation of the people in the photos is confounded with physical qualities of the photos (e.g., comparing criminal mug shots to the "NimStim" database, Valla, Ceci, & Williams, 2011). A second question is whether the images might have been selected to convey particular messages (e.g., images posted on dating websites might be selected for uploading because of the information they signal). In the current study, similar to others on accuracy from facial inference (Jones et al., 2012; Kramer & Ward, 2010; Little & Perrett, 2007; Penton-Voak et al., 2006), all images were generated by experimenters using a controlled environment and through asking the targets to provide a neutral expression for the camera, with controllable cues relating to hair, clothing, jewlery, and cosmetics minimised. This procedure

avoids the most obvious criticisms of selection bias. However, it does not mean that these images represent the ground truth about appearance. Having one's picture taken is a form of social interaction, and one could speculate that even this rather minimal context might still have induced small postural or expression differences in the photo targets, which were correlated with their true trait levels. But, even if this speculation were correct -- and at present we have no evidence one way or the other -- it would mean that accuracy arises from the social signals sent by the people in the photos, in the absence of any obvious social goal (such as appearing attractive for a dating website).

This leads us to the second point raised by Todorov and Porter (2014): what are we to make of accurate social inference from appearance in any case, given that different images of the same person can lead to different inferences? Our perspective is that yes, from the universe of possible images that could be taken of a person's face, many different impressions can be made: some will appear warm, some will be frightening, some will look confused, some will be unflattering. Some will reflect true traits, others transient emotional states, and others will be completely misleading. What we and others have shown is that it is relatively easy to isolate - from the universe of possible face images - a rather ordinary and unremarkable context that can, on average, reveal something about people's trait levels from mere appearance.

References

Alexander, R. D. (1990). *How did humans evolve? Reflections on the uniquely unique species*. University of Michigan: Ann Arbor.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would Jesse Jackson 'fail' the Implicit Association Test?" *Psychological Inquiry*, *15*(4), 257–278.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948–17953. https://doi.org/10.1073/pnas.0705435104

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence on social perception and cognition. In J. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York, NY: Guilford.

Biesanz, J. C. (2010). The Social Accuracy Model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, *45*(5), 853–885. https://doi.org/10.1080/00273171.2010.519262

Blanton, H., Jaccard, J., Christie, C., & Gonzales, P. M. (2007). Plausible assumptions, questionable assumptions and post hoc rationalizations: Will the real IAT, please stand up? *Journal of Experimental Social Psychology*, *43*(3), 399–409. https://doi.org/10.1016/j.jesp.2006.10.019

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

Bonnefon, J.-F., Hopfensitz, A., & Neys, W. D. (2017). Trustworthiness perception at zero acquaintance: Consensus, accuracy, and prejudice. *The Behavioral and Brain Sciences*, *40*, e4. https://doi.org/10.1017/S0140525X15002319

Boothroyd, L. G., Cross, C. P., Gray, A. W., Coombes, C., & Gregson-Curtis, K. (2011). Perceiving the facial correlates of sociosexuality: Further evidence. *Personality and Individual Differences*, *50*(3), 422–425. https://doi.org/10.1016/j.paid.2010.10.017

Boothroyd, L. G., Jones, B. C., Burt, D. M., DeBruine, L. M., & Perrett, D. I. (2008). Facial correlates of sociosexuality. *Evolution and Human Behavior*, *29*(3), 211–218. https://doi.org/10.1016/j.evolhumbehav.2007.12.009

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. https://doi.org/10.1016/j.jrp.2009.03.007

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science*, *25*(5), 1132–1139. https://doi.org/10.1177/0956797614523297

Daros, A. R., Ruocco, A. C., & Rule, N. O. (2016). Identifying mental disorder from the faces of women with borderline personality disorder. *Journal of Nonverbal Behavior*, *40*(4), 255–281. https://doi.org/10.1007/s10919-016-0237-9

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*(3), 347–368. https://doi.org/10.1037/a0014211

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290. https://doi.org/10.1037/h0033731

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Garwood, S. G., Cox, L., Kaplan, V., Wasserman, N., & Sulzer, J. L. (1980). Beauty is only “name” deep: The effect of first-name on ratings of physical attraction. *Journal of Applied Social Psychology*, *10*(5), 431–435. https://doi.org/10.1111/j.1559-1816.1980.tb00721.x

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us?: Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*(2), 181–193. https://doi.org/10.1111/j.1745-6916.2007.00036.x

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of Implicit Attitudes. *Journal of Experimental Psychology: General*, *143*(3), 1369–1392. https://doi.org/10.1037/a0035028

Joiner, T., & Coyne, J. C. (1999). *The interactional nature of depression: Advances in interpersonal approaches*. Washington, DC, US: American Psychological Association. https://doi.org/10.1037/10311-000

Jones, A. L. (2018). The influence of shape and colour cue classes on facial health perception. *Evolution and Human Behavior*, *39*(1), 19–29. https://doi.org/10.1016/j.evolhumbehav.2017.09.005

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: the contributions of facial shape, skin texture, and viewing angle. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(6), 1353–1361. https://doi.org/10.1037/a0027078

Kahneman, D., & Tresiman. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davis (Eds.), *Varieties of attention* (pp. 29–61). Orlando: Academic Press. Retrieved from http://www.scirp.org/(S(vtj3fa45qm1ean45vvffcz55))/reference/ReferencesPapers.aspx?ReferenceID=255111

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin*, *116*(2), 245–258. https://doi.org/10.1037/0033-2909.116.2.245

Kramer, R. S. S., & Jones, A. L. (2015). Do people's first names match their faces? *Journal of Articles in Support of the Null Hypothesis*, *12*(1).

Kramer, Robin S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *The Quarterly Journal of Experimental Psychology*, *63*(11), 2273–2287. https://doi.org/10.1080/17470211003770912

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (2nd ed., pp. 381–402). Oxford, UK: Blackwell Scientific Publications.

Leirer, V. O., Hamilton, D. L., & Carpenter, S. (1982). Common first names as cues for inferences about personality. *Personality and Social Psychology Bulletin*, *8*(4), 712–718. https://doi.org/10.1177/0146167282084018

Little, A. C. (2017). An evolutionary approach to accuracy in social perception. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X15002356

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, *28*(1), 18–27. https://doi.org/10.1016/j.evolhumbehav.2006.09.002

Little, A. C., & Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology*, *98*(1), 111–126.

Little, A. C., Třebický, V., Havlíček, J., Roberts, S. C., & Kleisner, K. (2015). Human perception of fighting ability: facial cues predict winners and losers in mixed martial arts fights. *Behavioral Ecology*, *26*(6), 1470–1475. https://doi.org/10.1093/beheco/arv089

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood. Physical attractiveness and accuracy in first impressions of personality. *Psychological Science*, *21*(12), 1777–1782. https://doi.org/10.1177/0956797610388048

Luxen, M. F., & Van De Vijver, F. J. R. (2006). Facial attractiveness, sexual selection, and personnel selection: When evolved preferences matter. *Journal of Organizational Behavior*, *27*(2), 241–255. https://doi.org/10.1002/job.357

Maynard Smith, J., & Harper, D. (2003). *Animal Signals*. OUP Oxford.

Mazur, A. (2005). *Biosociology of dominance and deference*. Maryland: Rowman & Littlefield Publishers.

Mazur, A., Mazur, J., & Keating, C. (1984). Military rank attainment of a West Point class: Effects of cadets' physical features. *American Journal of Sociology*, *90*(1), 125–150. https://doi.org/10.1086/228050

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442. https://doi.org/10.1006/jesp.2000.1470

McCrae, R. R., & Costa, P. T. (1989). The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of Personality and Social Psychology*, *56*(4), 586–595. https://doi.org/10.1037/0022-3514.56.4.586

Meier, B. P., Robinson, M. D., Carter, M. S., & Hinsz, V. B. (2010). Are sociable people more beautiful? A zero-acquaintance analysis of agreeableness, extraversion, and attractiveness. *Journal of Research in Personality*, *44*(2), 293–296. https://doi.org/10.1016/j.jrp.2010.02.002

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*(3), 455–469. https://doi.org/10.1037/0096-3445.132.3.455

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. https://doi.org/10.1177/0146167209346309

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity. *Personality and Social Psychology Bulletin*, *31*(2), 166–180. https://doi.org/10.1177/0146167204271418

Nosek, B. A., Ranganath, K. A., Smith, C. T., Chugh, D., Olson, K. R., Lindner, N. M., … Hansen, J. J. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88. https://doi.org/10.1080/10463280701489053

Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans prefer Republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. *Social Psychological and Personality Science*, *3*(5), 605–613. https://doi.org/10.1177/1948550611432770

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. https://doi.org/10.1007/s10919-009-0082-1

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion (Washington, D.C.)*, *5*(4), 498–502. https://doi.org/10.1037/1528-3542.5.4.498

Olson, M. A., Fazio, R. H., & Hermann, A. D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science*, *18*(4), 287–291. https://doi.org/10.1111/j.1467-9280.2007.01890.x

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition*, *24*(5), 607–640. https://doi.org/10.1521/soco.2006.24.5.607

Pound, N., Penton-Voak, I. S., & Brown, W. M. (2007). Facial symmetry is positively associated with self-reported extraversion. *Personality and Individual Differences*, *43*(6), 1572–1582. https://doi.org/10.1016/j.paid.2007.04.014

Quist, M. C., Watkins, C. D., Smith, F. G., DeBruine, L. M., & Jones, B. C. (2011). Facial masculinity is a cue to women's dominance. *Personality and Individual Differences*, *50*(7), 1089–1093. https://doi.org/10.1016/j.paid.2011.01.032

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. https://doi.org/10.1016/j.jesp.2007.12.001

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PLoS One*, *5*(1), e8733.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology*, *45*(6), 1245–1251. https://doi.org/10.1016/j.jesp.2009.07.010

Scott, N. J., Kramer, R. S. S., Jones, A. L., & Ward, R. (2013). Facial cues to depressive symptoms and their associated personality attributions. *Psychiatry Research*, *208*(1), 47–53. https://doi.org/10.1016/j.psychres.2013.02.027

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology*, *31*(3), 410–414.

Siibak, A. (2009). Constructing the self through the photo selection - Visual impression management on social networking websites. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *3*(1). Retrieved from https://cyberpsychology.eu/article/view/4218

Smith-McLallen, A., Johnson, B. T., Dovidio, J. F., & Pearson, A. R. (2006). Black and white: The role of color bias in implicit race bias. *Social Cognition*, *24*(1), 46–73. https://doi.org/10.1521/soco.2006.24.1.46

Sparks, A., Burleigh, T., & Barclay, P. (2016). We can see inside: Accurate prediction of Prisoner's Dilemma decisions in announced games following a face-to-face interaction. *Evolution and Human Behavior*, *37*(3), 210–216. https://doi.org/10.1016/j.evolhumbehav.2015.11.003

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354. https://doi.org/10.1177/0956797610362647

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, *25*(7), 1404–1417. https://doi.org/10.1177/0956797614532474

Tognetti, A., Berticat, C., Raymond, M., & Faurie, C. (2013). Is cooperativeness readable in static facial features? An inter-cultural approach. *Evolution and Human Behavior*, *34*(6), 427–432. https://doi.org/10.1016/j.evolhumbehav.2013.08.002

Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, *907*, 114–131.

Uhlmann, E. L., Poehlman, T. A., & Nosek, B. A. (2012). Automatic associations: Personal attitudes or cultural knowledge? In *Ideology, psychology, and law* (pp. 228–260). New York, NY, US: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737512.003.0009

Uleman, J. S. (1987). Consciousness and control: The case of spontaneous trait inferences. *Personality and Social Psychology Bulletin*, *13*(3), 337–354. https://doi.org/10.1177/0146167287133004

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360. https://doi.org/10.1146/annurev.psych.59.103006.093707

Valla, J. M., Ceci, S. J., & Williams, W. M. (2011). The accuracy of inferences about criminality based on facial appearance. *Journal of Social, Evolutionary, and Cultural Psychology*, *5*(1), 66.

van Leeuwen, M. L., & Macrae, C. N. (2004). Is beautiful always good? Implicit benefits of facial attractiveness. *Social Cognition*, *22*(6), 637–649. https://doi.org/10.1521/soco.22.6.637.54819

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

Ward, R., & Scott, N. J. (2018). Cues to mental health from men's facial appearance. *Journal of Research in Personality*, *75*, 26–36. https://doi.org/10.1016/j.jrp.2018.04.007

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

Young, R. K., Kennedy, A. H., Newhouse, A., Browne, P., & Thiessen, D. (1993). The effects of names on perception of intelligence, popularity, and competence. *Journal of Applied Social Psychology*, *23*(21), 1770–1788. https://doi.org/10.1111/j.1559-1816.1993.tb01065.x

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*(6), 603–623. https://doi.org/10.1007/BF01065855

Zwebner, Y., Sellier, A.-L., Rosenfeld, N., Goldenberg, J., & Mayo, R. (2017). We look like our names: The manifestation of name stereotypes in facial appearance. *Journal of Personality and Social Psychology*, *112*(4), 527–554. https://doi.org/10.1037/pspa0000076