21.08.17. Facial First Impressions of Partner Preference Traits:

Trustworthiness, Status, and Attractiveness

Jennifer K. South Palomares and Andrew W. Young

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

Correspondence should be addressed to Jennifer K. South Palomares, Department of Education, University of York, York YO10 5DD, UK, jennifer.south.palomares@york.ac.uk

Abstract

This research used the minimal exposure paradigm to examine facial first impressions of traits of trustworthiness, status, and attractiveness, considered important in verbal models of partner preferences. Heterosexual participants rated opposite-sex faces comprising either naturalistic images or youthful-looking averaged faces on trustworthiness, status, and attractiveness following 33ms, 100ms, and 500ms masked presentation. The pattern masks were phase-scrambled to provide the same overall colour composition, brightness, and spatial frequency content as the presented faces. Trustworthiness, status, and attractiveness judgements were all reliable at above-chance levels even at 33ms presentation, and extra time (100ms or 500ms) only led to modest improvement in the correspondence with an independent set of time-unconstrained judgements. The increasing prevalence of online images and internet-based relationships make these findings timely and important.

Theories of verbally expressed romantic partner preferences emphasise three factors: warmth-trustworthiness, status-resources, and vitality-attractiveness (Fletcher, Simpson, Thomas, & Giles, 1999), which have been validated in numerous questionnaire-based studies (Campbell, Simpson, Kashy, & Fletcher, 2001; Fletcher, Kerr, Li, & Valentine, 2014; Fletcher, Simpson, & Thomas, 2000a, 2000b; Fletcher et al., 1999). Faces are a salient source of information during relationship initiation (e.g., online dating). For instance, the online dating company Meetic Group moderates over 15 billion profiles and claims to be at the origin of six million European couples (Meetic Group, 2016). Face photographs are nearly always salient in these profiles and, as users often search quickly through many profiles, it is important to understand what information they might get from a quick glance at a face. Here, our interest is in whether verbally expressed partner preference traits can be evaluated from facial first impressions.

Various traits based on facial appearance can be evaluated from brief presentations, but this has not been investigated in the context of evaluations underlying partner preferences or with everyday faces. Our aim was to explore trait evaluations from a leading model of verbal partner preferences following time-limited presentation of faces. The selection of traits for this research followed from models of relationships but, to avoid biasing participants' evaluations toward partner preferences, there was no mention of relationships in the paradigm.

Consistent with suggestions that face evaluations are fast and automatic (Todorov et al., 2005), research has shown that such judgments can be made at above-chance levels following minimal presentation times (Ballew & Todorov, 2007; Bar, Neta, & Linz, 2006; Locher, Unger, Sociedade, and Wahl, 1993; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Here, we are interested in how well impressions formed to brief supraliminal presentations correspond to those made with unlimited viewing time. For example, Todorov

et al. (2009) found that following 33ms presentation time evaluations of facial trustworthiness corresponded somewhat with independent time-unconstrained judgements, and this correspondence improved with increased presentation times (up to 167ms).

The current research examined the relative salience of traits deemed important in verbal models of partner preferences using the minimal exposure paradigm, firstly with naturalistic faces and then with youthful-looking averaged faces. We focused on judgments of trustworthiness, status, and attractiveness because these correspond to the factors identified by Fletcher et al.'s (1999) leading verbal model of partner preferences and because similar traits are important in models of facial first impressions (Sutherland et al., 2013). Study 1 was used to create a set of consensual judgements of trustworthiness, status, and attractiveness from everyday faces when given unlimited presentation time. These judgements formed an independent standard to which performance under the time-limited presentation conditions of Study 2 could be compared. Study 2 involved ratings of trustworthiness, status, or attractiveness with time-constrained (33ms-500ms) presentation of each face. To prevent further visual processing of each image, it was immediately followed by a pattern mask comprising a Fourier phase-scrambled version of the same image. Because our interest was in first impressions, participants viewed each face once in Studies 1-2. Studies 3-4 examined trait inferences to youthful-looking averaged faces that more closely matched the age of our participants and included a gender discrimination task to establish how well participants could see the faces at short presentation times.

Study 1

Participants rated naturalistic faces on trustworthiness, status, and attractiveness following unlimited presentation time.

Method

Participants. Fourteen participants, university students, were recruited via the

University of York (50% male, mean age of 21 years, *SD*=2.88). Participants were self-reported native English speakers, raised in a Western environment. The participants provided written consent to procedures approved by the Ethics Committee of the University of York Psychology Department. No data from participants were excluded. Participants did not take part in the other reported studies.

Face images. Studies typically standardise faces, removing between-image differences considered unwanted "noise". Our approach is novel, focusing on impressions to naturalistic, everyday face photographs (termed ambient images by Jenkins, White, Van Montfort, & Burton, 2011), including all variability (e.g., pose) this entails. Image differences can have a pronounced impact on facial impressions (Burton, Jenkins, & Schweinberger, 2011; Jenkins et al., 2011; Sutherland, Young, & Rhodes, 2017; Todorov & Porter, 2014).

The study used a database of 1,000 face images (50% male, Santos & Young, 2005, 2008, 2011), representing Caucasian non-famous adults. Like other face databases (e.g., Oosterhof & Todorov, 2008), to avoid other-race effects (Anzures et al., 2013; Feng et al., 2011; Hugenberg & Bodenhausen, 2003; O'Toole, Natu, & Toole, 2013), non-Caucasian faces were not included. The images were taken from the internet and were unconstrained in terms of variability (e.g., expression, age, amongst others). Everything except Caucasian adult appearance was unstandardised. Images were resized to 150 pixels in height and cropped to reveal the individuals' head and shoulders. See Figure 1 for examples.
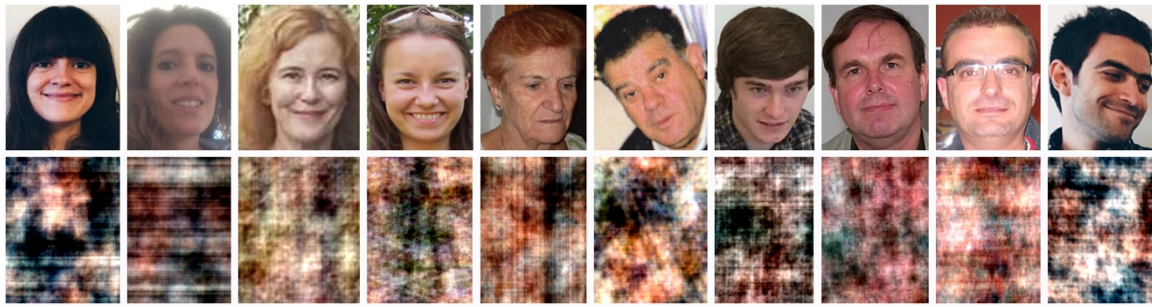
Figure 1. Example ambient images like those used in the study, from the authors' personal collections (upper panel). Photographs from the database are not shown for copyright reasons. The lower panel shows corresponding Fourier phase-scrambled masks, as used in Studies 2 and 4.

Traits. Ratings were already available for trustworthiness (*n*=20; Sutherland et al., 2013), and attractiveness (*n*=6; Santos & Young, 2005, 2008, 2011). Ratings on status and additional ratings on attractiveness were collected using the same method (to arrive at *n*=10 per trait).

Procedure. Participants were informed that the study involved facial evaluations and that the task was self-paced, but to rely on their first impressions (Sutherland et al., 2013; Todorov, Mandisodza, Goren, & Hall, 2005). Participants completed six practice trials, rating faces randomly selected from the database, and then rated all 1,000 images, in a random order on one trait, to avoid carry-over effects (e.g., Hamermesh & Abrevaya, 2013; Rhodes, 2006). Ten participants (five male) were randomly assigned to rate status and four participants (two male) were randomly assigned to rate attractiveness. Ratings were made on a Likert scale (e.g., 1: very unattractive – 7: very attractive). Images remained on the screen while participants made their judgement. The inter-trial interval was 750ms. On completion, participants were debriefed and reimbursed with a small payment. The task was programmed using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, USA) and took 60 minutes.

Results and Discussion

Reliability of trait judgements across raters was good (Nunnally & Bernstein, 1994) and in accordance with previous studies (e.g. Sutherland et al., 2013), as presented in Table 1. We do not maintain that these evaluations are valid, though, what matters is that the time-unconstrained ratings were largely consensual (reliable) across participants. The data of interest involved the mean rating for each face on each trait. The Study 1 mean trait ratings for each face formed an independent point of comparison to the time-constrained ratings in Study 2.

Table 1. Range of mean ratings of ambient images on trustworthiness, status, and attractiveness following time-unlimited presentation, from Study 1. Cronbach's alphas of trait ratings are also shown.

Trait	Mean range	Cronbach's alpha
Trustworthiness	1.80-6.05	.87
Status	1.70-6.00	.81
Attractiveness	1.50-6.10	.87

Study 2

To examine first impressions to brief presentation, participants rated naturalistic faces on trustworthiness, status, and attractiveness following 33ms-500ms masked presentation.

Method

Participants. Thirty participants, university students, were recruited via the University of York (50% male, mean age 21 years, *SD*=3.09). Participants were self-reported heterosexual native English speakers, raised in a Western environment. The participants provided written consent to procedures approved by the Ethics Committee of the University of York Psychology Department. No data from participants were excluded. Participants did not take part in the other reported studies. Post-hoc power analysis confirmed that the sample size was adequate to achieve over .80 power when examining whether trait evaluations could demonstrate above-chance reliability.

Face images. The same database described for Study 1. From this database, 300 images of male faces were randomly selected and grouped into ten sets of 30 images, and 300 randomly selected female faces were likewise grouped into ten sets of 30 images. No image was repeated across these sets of stimuli.

Masks. Image processing was performed in MATLAB R2015b to create a Fourier phase-scrambled mask per image (see Figure 1). Fourier-scrambling was achieved by adding the same random phase structure to the existing three (rgb) phase structures within the original face. Hence, the overall colour composition, brightness, and spatial frequency of each scrambled mask was the same as in the original face (cf. Baseler, Harris, Young, & Andrews, 2014).

Procedure. Participants were informed that the study involved facial evaluations and that the task was self-paced, but to rely on their first impressions (Sutherland et al., 2013; Todorov et al., 2005). Participants completed the minimal exposure time task, rating opposite-sex faces on trustworthiness, status, and attractiveness following three different presentation times (33ms, 100ms, and 500ms). There were three main blocks of trials, corresponding to the three trait ratings. Within each main trait rating block there were three subsidiary blocks with different presentation times (33ms, 100ms, and 500ms). Trait order

and presentation time order were counterbalanced across participants.

Within each main trait rating block, participants rated 30 practice stimuli before the main blocks of 30 trials with 33ms, 100ms, and 500ms presentation times. Each trial started with a fixation cross presented for 500ms at the centre of the screen. After viewing each face for a given presentation time, a mask was presented to overlap the previous image and this remained on the screen until participants rated the original image on trustworthiness (1: very untrustworthy – 7: very trustworthy), status (1: low status – 7: high status), or attractiveness (1: very unattractive – 7: very attractive). The inter-trial interval was 750ms. All participants rated the same set of 30 practice images and the remaining nine sets of 30 faces were rotated randomly around the conditions (three presentation times and three rated traits). Participants did not rate any face twice. On completion, participants were debriefed and reimbursed with a small payment. Stimuli were presented on an LCD screen and the task was programmed using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, USA) and took 25 minutes.

Results and Discussion

Reliability of trait judgements across raters could not be examined as the counterbalancing meant that few participants rated each set of faces following a given presentation time. To measure performance, we therefore correlated each participant's trait ratings for each face at 33ms, 100ms, and 500ms with the mean trait ratings for the independent time-unconstrained judgements of the same images from Study 1. Our measure of performance thus involved quantifying the agreement between an individual participant's ratings (at different presentation times) with consensual judgements of the same images from Study 1. These correlations were transformed using Fisher's z (Fisher, 1915) for statistical analysis.

A key question concerns whether traits can be evaluated at above-chance levels at the different presentation times. One-sample t-tests comparing the Fisher's z scores of the

conditions to zero revealed that the correlations for each condition were significant (see Table

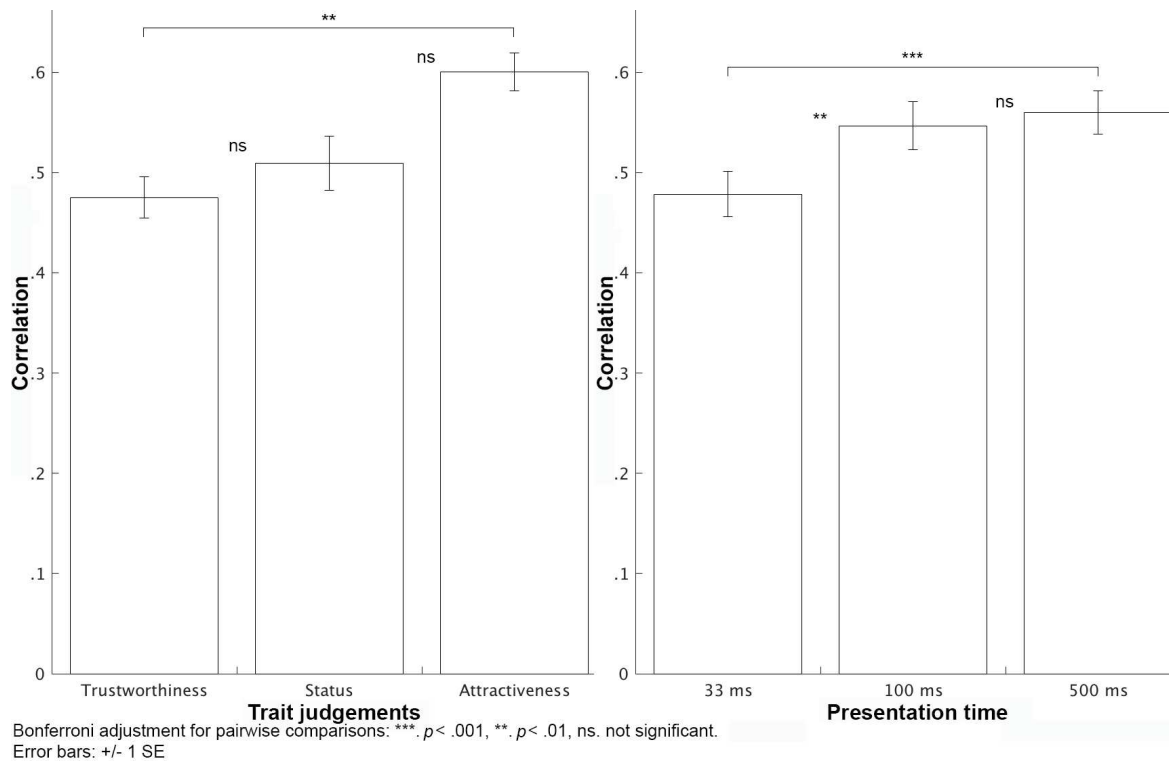2), showing above-chance performance in all conditions.

Table 2. One-sample t-tests comparing Fisher z scores representing the relationship between time-constrained face evaluations from Study 2 and an independent set of time-unlimited face judgements (derived from Study 1) to zero. All t-tests were significant at p<.001.

Trait	Presentation time	t-test
Trustworthiness	33ms	$t(29) = 11.19$
	100ms	$t(29) = 12.31$
	500ms	$t(29) = 11.04$
Status	33ms	$t(29) = 8.34$
	100ms	$t(29) = 9.04$
	500ms	$t(29) = 11.51$
Attractiveness	33ms	$t(29) = 14.56$
	100ms	$t(29) = 14.79$
	500ms	$t(29) = 16.08$

A more nuanced approach is to ask whether performance improved across

presentation time. Hence, a repeated-measures 3 x 3 ANOVA was conducted between traits

(trustworthiness, status, and attractiveness) and presentation times (33ms, 100ms, and

500ms). The model did not violate sphericity. There were significant main effects of traits:

$F(2, 58) = 6.89, MSE = 0.81, p = .002, \eta_p^2 = .19$ and presentation times: $F(2, 58) = $

$10.95, MSE = 0.42, p < .001, \eta_p^2 = .27$. The interaction between presentation times and

traits did not reach significance: $F(4, 116) = 1.28, MSE = 0.07, p = .282, \eta_p^2 = .04$.

The left panel of Figure 2 shows the main effect of traits. Examination using

Bonferroni's adjustment revealed that attractiveness was more reliably detected from ambient

images than was trustworthiness (*p*<.01); therefore, rated attractiveness was more consensual

across participants than rated trustworthiness. The right panel of Figure 2 shows the main

effect of presentation times. Findings revealed a modest improvement in the correspondence

with the independent set of time-unconstrained judgements following 100ms or 500ms

presentation times, relative to 33ms ($ps<.01$). Further, relative to 100ms presentation time,

evaluations made following 500ms did not result in better correspondence with the

independent time-unconstrained judgements.

Bonferroni adjustment for pairwise comparisons: ***. $p<.001$, **. $p<.01$, ns. not significant.
Error bars: +/- 1 SE

Figure 2. The main effects of traits (left panel) and presentation times (right panel) in Study

2. Mean correlations (and SE) between ratings of trustworthiness, status, and attractiveness

of ambient faces presented for limited time (33ms to 500ms) in Study 2 and an independent

set of time-unconstrained judgements (derived from Study 1).

From these findings, it is clear that seeing a face for 100ms suffices to make a reliable

first impression, and that even a presentation as short as 33ms creates only a small decrement.

It is possible, though, that the less correlated judgments following 33ms presentation might

simply reflect greater noise if participants missed seeing stimuli on some trials. Hence, Study

4 incorporated an additional gender categorisation task to determine whether participants

could detect a highly salient facial characteristic (gender) following 33ms and 100ms

presentation.

Studies 1-2 used everyday images of adult faces unconstrained on age, whereas our participants were young adults. As age is a correlate of perceived trustworthiness, status and attractiveness (Sutherland et al., 2013; Todorov, 2017), we explored whether these traits could also be reliably determined from brief presentations when the range of ages was constrained to be relatively close to the ages of the participants. To achieve this, we used image averaging techniques (Sutherland, Rhodes, & Young, 2017) to create youthful-looking face-like averaged faces high or low on trustworthiness, status, and attractiveness for Studies 3-4.

Study 3

Sutherland et al. (2013) showed that averaging ambient images can create prototypes conserving the consistent cues underlying certain traits. We used the same method to create averages of youthful-looking ambient images based on trustworthiness, status, and attractiveness ratings, which were evaluated by a new group of participants in Study 3 following unlimited presentation time. These judgements formed an independent standard to which performance under the time-constrained presentation conditions of Study 4 could be compared.

Method

Participants. Twenty participants, university students, were recruited via the University of York (50% male, mean age of 21 years, SD =3.91). Participants were self-reported heterosexual native English speakers, raised in a Western environment. The participants provided written consent to procedures approved by the Ethics Committee of the University of York Psychology Department. No data from participants were excluded. Participants did not take part in the other reported studies.

Face-like images. We followed Sutherland et al.'s (2013) image averaging procedure

using PsychoMorph (Tiddeman, Burt, & Perrett, 2001) to create 120 face-like averaged

images from the ambient face database described for Study 1. These averaged images (see

Figure 3) represented high and low trustworthiness, status, and attractiveness. Each image

was made by averaging six high and six low rated images, using only images representing

younger adults (based on age ratings from Santos & Young, 2005, 2008, 2011).
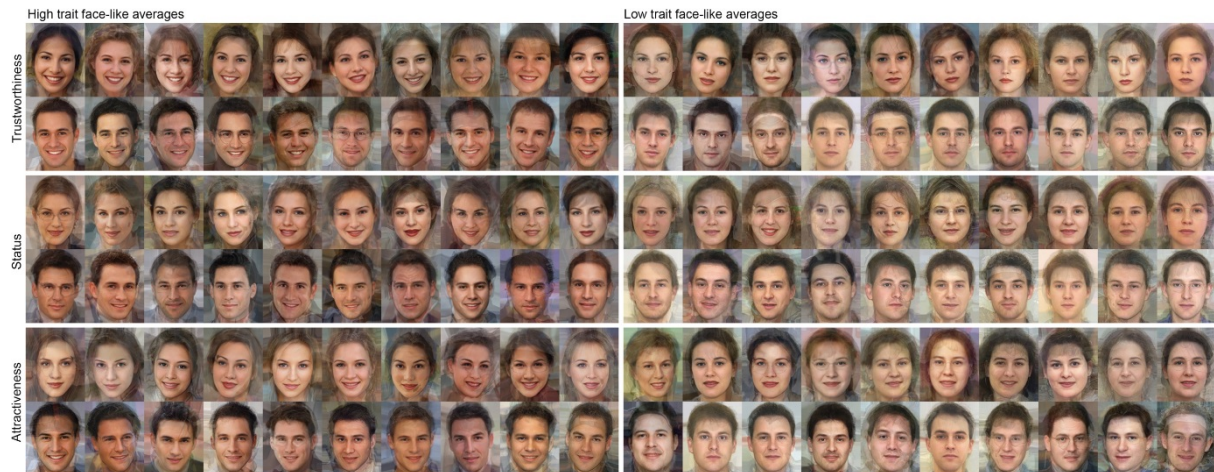
Figure 3. Stimuli used in Studies 3 and 4. These are averaged youthful-looking faces. Each average is made from six ambient images representing high (left panels) and low (right panels) trustworthiness (upper rows), status (middle rows), and attractiveness (lower rows). Averages created from female images are shown in the top panels and male images in the bottom panels.

Procedure. Participants rated all 120 images on trustworthiness, status, and attractiveness in a blocked design; trait order was counterbalanced. The task took 25 minutes. All other procedural details were the same as for Study 1.

Results and discussion

Reliability of trait judgements across raters was good (Nunnally & Bernstein, 1994) and in line with previous studies (e.g. Sutherland et al., 2013), as presented in Table 3. The data of interest involved the mean rating for each face per trait (see Table 3).

Table 3. Mean and SD of ratings of male and female youthful-looking averaged face images (representing high and low trustworthiness, status, and attractiveness) on trustworthiness, status, and attractiveness following time-unlimited presentation, from Study 3. Cronbach's alphas of trait ratings are also shown.

Trait ratings	Face gender	Faces	Mean	*SD*	Cronbach's alpha
Trustworthiness	Male	High trustworthiness	4.15	.36	.90
		Low trustworthiness	3.04	.26	
	Female	High trustworthiness	5.17	.35	
		Low trustworthiness	3.55	.30	
Status	Male	High status	5.00	.45	.90
		Low status	3.31	.40	
	Female	High status	4.65	.43	
		Low status	3.57	.46	
Attractiveness	Male	High attractiveness	4.40	.55	.96
		Low attractiveness	2.48	.42	
	Female	High attractiveness	5.04	.35	
		Low attractiveness	3.04	.58	

To further test the validity of the averaged stimuli, the ratings were subjected to a repeated-measures ANOVA for stimulus type (high/low male/female attractiveness, high/low male/female status, and high/low male/female trustworthiness). The model did not violate sphericity. The ANOVA revealed a significant effect of stimulus type: $F(11, 99) = 43.83, MSE = 8.20, p < .001, \eta_p^2 = .83$. Examination of this effect using Bonferroni's adjustment revealed significant differences (*p*s<.01) between high and low faces manipulated to represent each trait, separated by face gender. Female and male images manipulated to represent high and low levels of each trait were evaluated in accordance with the manipulations (high trustworthiness faces were evaluated as being higher in trustworthiness, relative to low trustworthiness faces, and so on). Hence, these face-like averages were validated for use in Study 4 and these judgements formed an independent standard to compare to the time-constrained trait evaluations in Study 4.

Study 4

Participants rated the youthful-looking averaged images on trustworthiness, status, and attractiveness following 33ms-500ms presentation and completed a gender discrimination task to ascertain whether faces at 33ms and 100ms are visible to participants.

Method

Participants. Sixty-two participants, university students, were recruited via the University of York (43% male, mean age of 22 years, *SD* =4.27). Participants were self-reported heterosexual native English speakers, raised in a Western environment. The participants provided written consent to procedures approved by the Ethics Committee of the University of York Psychology Department. Data from one participant's trustworthiness 33ms and 100ms tasks were not included as he gave identical ratings throughout. Participants did not take part in the other reported studies. Post-hoc power analysis confirmed that the sample size was adequate to achieve over .80 power when examining whether trait evaluations could demonstrate above-chance reliability.

Face images. The study used the face-like images described for Study 3 in the minimal exposure time task. Ambient images (described in Study 1) rated high for gender typicality (high masculinity ratings of male faces, high femininity ratings of female faces; Sutherland et al., 2013) were used for the gender discrimination task.

Masks. As per Study 2, image processing was performed in MATLAB R2015b to create a Fourier phase-scrambled mask for each face.

Procedure. Participants were informed that the study involved facial evaluations and that the task was self-paced, but to rely on their first impressions (Sutherland et al., 2013; Todorov et al., 2005). Following informed consent, participants completed a gender discrimination task, then the minimal exposure time task, and finally another gender discrimination task. The gender discrimination task involved six practice trials followed by ten trials viewing faces presented for 33ms and ten trials viewing faces for 100ms in a

blocked design (exposure time was counterbalanced). Each face was immediately followed by a mask overlapping the previous image. The mask remained on the screen until participants responded to the question: *Is this face male or female?* (1: male or 7: female). The faces viewed by participants in the final gender discrimination task were different from those in the initial gender task.

Procedural details for the minimal exposure time task were the same as for Study 2, excepting that within each main trait rating block, participants rated 20 opposite sex faces on the trait they had been manipulated to represent (i.e., participants rated 20 faces on trustworthiness: 10 high and 10 low trustworthiness images). Participants rated the same faces on a given trait, in a random order, for each presentation time. Participants rated six practice stimuli before the main blocks of 20 trials viewing faces for 33ms, 20 trials at 100ms, and 20 trials at 500ms presentation. Stimuli were presented on an LCD screen and the tasks were programmed using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, USA) and took 25 minutes.

Results and discussion

In the gender discrimination task, all participants classified stimuli with at least 75% accuracy following 33ms-100ms presentation. One-sample *t*-tests revealed above-chance accuracy for 33ms and for 100ms, regardless of task order (i.e., whether participants completed the task before or after trait ratings; see Table 4).

Table 4. One-sample t-tests comparing the percent accuracy from the gender discrimination tasks from Study 4 to chance level (50% accuracy). All t-tests were significant at p<.001, demonstrating above-chance accuracy for faces presented at 33ms and 100ms regardless of whether participants completed the gender discrimination tasks before or after completing the trait rating task. Mean percent accuracy and SD are shown.

Presentation time	Task order	*t*-test	Mean	*SD*
33ms	Before trait ratings	*t*(61)=27.90	85%	9.88
	After trait ratings	*t*(61)=25.04	86%	11.36
	Overall	*t*(61)=43.27	86%	6.47
100ms	Before trait ratings	*t*(61)=51.91	96%	6.92
	After trait ratings	*t*(61)=45.88	94%	7.59
	Overall	*t*(61)=67.51	95%	5.24

To further examine accuracy in the gender discrimination task, a repeated-measures 3 x 3 ANOVA was conducted between presentation times (33ms and 100ms) and task order (before and after completing trait ratings). The model did not violate sphericity. There was a significant main effect for presentation times revealing that accuracy was better following 100ms (relative to 33ms) presentation: $F(1,61) = 104.27, MSE = 5425.81, p < .001, \eta_p^2 = .63$. Hence, it is possible that participants 'missed' some faces at 33ms. There was no significant main effect for task order: $F(1,61) = .02, MSE = 1.61, p = .904, \eta_p^2 = .00$ and no significant interaction between presentation times and task order: $F(1,61) = 1.24, MSE = 103.23, p = .271, \eta_p^2 = .02$. Note that the overall pattern of results reported below did not change when only considering data from individuals with at least 85% accuracy in the gender discrimination task.

For the trait ratings, because stimuli were repeated across conditions, we were able to demonstrate good reliabilities across raters (Nunnally & Bernstein, 1994) in line with previous studies (e.g. Sutherland et al., 2013), as shown in Table 5. To measure performance for trait ratings at each presentation time, we correlated each participant's trait ratings for

each face at 33ms, 100ms, and 500ms with the mean trait ratings for the independent time-unconstrained judgements of the same images from Study 3. These correlations were transformed using Fisher's z (Fisher, 1915) for statistical analysis. One-sample t-tests comparing the Fisher's z scores of the conditions to zero revealed that the correlations of each condition were significantly above-chance level (see Table 5).

Table 5. One-sample t-tests comparing Fisher z scores, representing the relationship between time-constrained face evaluations from Study 4 and an independent set of time-unconstrained face judgements (derived from Study 3), to zero. All t-tests were significant at p<.001. Cronbach's alphas of time-constrained trait ratings are also presented.

Trait	Presentation time	t-test	Cronbach's alpha	
			Male faces	Female faces
Trustworthiness	33ms	$t(60)=12.81$	.96	.94
	100ms	$t(60)=15.42$	.96	.96
	500ms	$t(61)=14.05$	.96	.93
Status	33ms	$t(61)=9.38$	.93	.79
	100ms	$t(61)=14.03$	.95	.90
	500ms	$t(61)=13.57$	.94	.88
Attractiveness	33ms	$t(61)=17.54$	.96	.95
	100ms	$t(61)=24.92$	.98	.98
	500ms	$t(61)=33.31$	.98	.98

To examine Fisher's z score differences between conditions, a repeated-measures 3 x 3 ANOVA was conducted between traits (trustworthiness, status, and attractiveness) and presentation times (33ms, 100ms, and 500ms). A Greenhouse-Geisser correction was used as the model violated sphericity. There was a significant main effect for traits: $F(1.79, 107.17) = 32.66, MSE = 7.51, p < .001, \eta_p^2 = .35$ and for presentation times: $F(1.98, 118.51) = 18.69, MSE = 1.41, p < .001, \eta_p^2 = .24$. However, there was also a significant interaction between presentation times and traits: $F(3.68, 220.90) = 10.30, MSE = 0.58, p < .001, \eta_p^2 = .15$.

This interaction effect is presented in Figure 4. Decomposition of the interaction using Bonferroni's adjustment revealed that the reliability of trustworthiness judgements was not affected by presentation time, whereas judgements of status and attractiveness were more reliable at 100ms and 500ms presentation. Similar to Study 2, attractiveness was more reliably detected than trustworthiness ($p<.01$) or status ($p<.001$); in other words, rated attractiveness was more consensual across participants than rated trustworthiness or status. Trustworthiness was also more reliably detected than status ($p<.01$). As for Study 2, there was a modest improvement in the correspondence with the independent set of time-unconstrained judgements following 100ms or 500ms presentation times, relative to 33ms (ps$<.001$). Further, relative to 100ms presentation time, evaluations made following 500ms did not result in significantly better correspondence with the independent time-unconstrained judgements.
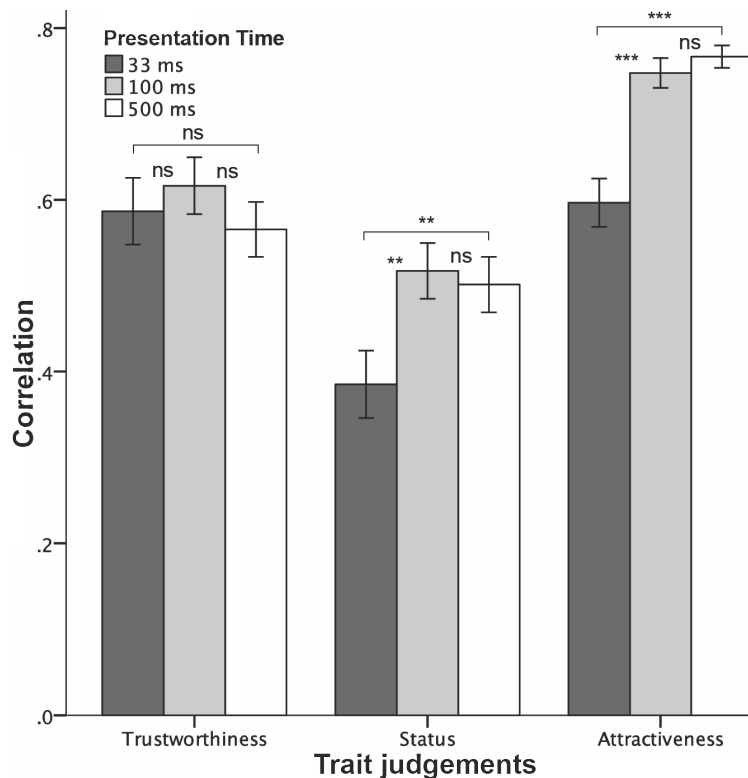
Bonferroni correction for pairwise comparisons:
***. $p < .001$, **. $p < .01$, ns. not significant
Error bars: +/- 1 SE

Figure 4. The interaction of traits and presentation times in Study 4. Mean correlations (and SE) between ratings of trustworthiness, status, and attractiveness of youthful-looking averaged faces presented for limited time (33ms to 500ms) in Study 4 and the independent set of time-unconstrained judgements (derived from Study 3).

General Discussion

Our studies investigated trait evaluations related to a leading model of verbal partner preferences from time-limited masked presentations of face stimuli. We cross-validated the findings using naturally occurring, ambient, face images (Study 2) and relatively youthful-looking face-like averaged images (Study 4). The reliability of first impressions of trustworthiness, status, and attractiveness following 33ms, 100ms, and 500ms presentation was examined by correlating these with time-unlimited evaluations of the same faces on the same traits by a different group of participants.

Time-constrained trait evaluations showed above-chance reliability across all conditions, and we never found any differences between performance to 100ms and 500ms presentations. Given the careful masking of each stimulus with an individually configured Fourier mask, we can conclude that a single glance of 100ms is sufficient to form a first impression that will, on average, approximate that of other viewers. Thus, as per previous research, time-constrained judgements are not necessarily less reliable (Ballew & Todorov, 2007).

There were slight differences in our findings with 33ms presentation time. Although likely close to threshold, this still allowed supraliminal perception; none of our participants reported being unable to see faces at the 33ms presentation time. In Study 2 (with ambient images) reliability was slightly reduced across all judgements at 33ms, whereas in Study 4 (with youthful-looking averaged images) the reduction in reliability at 33ms was evident for ratings of status and attractiveness, but not for trustworthiness. We note, though, that evaluations of face gender were also less accurate at 33ms than 100ms in Study 4, raising the possibility that participants 'missed' seeing some of the faces. Although potentially interesting, however, this difference between 33ms and 100ms is minor; the key point is the ease with which consensual evaluations were reached to brief presentations. This finding is notable as such brief presentation times are not considered long enough for saccadic eye movements, curtailing visual exploration and underscoring the effectiveness of a "single glance" (Olivola & Todorov, 2010).

The present research focused on judgments of trustworthiness, status, and attractiveness because these represent the factors identified by a leading verbal model of partner preferences which is based around factors of warmth-trustworthiness, status-resources, and vitality-attractiveness (Fletcher et al., 1999). However, these verbal partner preference factors somewhat parallel factors identified in other studies of facial first

impressions: trustworthiness/valence, youthful-attractiveness, and dominance/competence

(Oosterhof & Todorov, 2008; Sutherland et al., 2013; Sutherland, Oldmeadow, & Young,

2016). Hence, the traits examined here are likely good approximations of the key components

underlying person perception in both romantic and non-romantic contexts.

Facial judgments are consequential, for instance, predicting government election

results and influencing romantic preferences (Ballew & Todorov, 2007; Fiore, Taylor,

Mendelsohn, & Hearst, 2008; Fletcher et al., 2014; Hall, Goren, Chaiken, & Todorov, 2009;

Hancock & Toma, 2009; Todd, Penke, Fasolo, & Lenton, 2007; Todorov, 2017). Thus,

identifying the salience of traits involved in romantic preferences is both important and

timely. From an evolutionary perspective, evaluations of trustworthiness and attractiveness

may be of particular significance for two reasons. Firstly, detection of trustworthiness is

likely to be linked to survival (Cosmides & Tooby, 1992). For instance, people may avoid

potentially harmful individuals based on first impressions of facial trustworthiness (Oosterhof

& Todorov, 2008). Secondly, researchers suggest that attractiveness signals fertility and

resistance to environmental and genetic stressors (Jasienska et al., 2006; Lassek & Gaulin,

2008; Møller, 1999; see Rhodes, 2006, for a review). Attractiveness remains salient out of the

lab, for example, in speed-dating paradigms (see Walster, Aronson, Abrahams, and

Rottman's, 1966, famous "Computer Dance" study). Hence, evaluations of trustworthiness,

status, and attractiveness may be relevant in approach behaviours within and beyond

romantic contexts.

Whilst the evolutionary perspective has become dominant in attractiveness research in

particular, studies have shown distinct contributions of environmental and genetic influences

(Germine et al., 2015). More generally, it is clear that some aspects of first impressions are

consensual (i.e. shared across participants) whilst others reflect more idiosyncratic

evaluations (Hönekopp, 2006). In this respect, our technique of correlating an individual

participant's ratings with the average ratings across an independent group of observers estimates the consensual component. Of note, in Studies 2 and 4 we found attractiveness judgements elicited the highest consensus (i.e., the largest correlations). Moreover, the cues subserving facial impressions interact in complex ways (Santos & Young, 2011; Sutherland et al., 2013; Todorov, 2017; Vernon, Sutherland, Young, & Hartley, 2014). For example, averaged high warmth-trustworthiness images depict smiling individuals, but smiling is not an exclusive cue to warmth-trustworthiness. Instead, what is important is possibly the type of smile, and certainly the way smiling is combined with other cues such as skin tone, age, and face shape. These cues are interdependent, allowing reliable evaluations to be made even when the impact of the powerful age cue was reduced in Studies 3-4.

Our research applied the minimal exposure paradigm to novel stimuli: ambient images and computer-averaged face-like images. Naturalistic ambient faces capture the cues involved in trait evaluations that might be absent in standardised images and form a key element of data-driven approaches (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). The ambient image approach provides a useful degree of ecological validity by shifting the emphasis of facial impressions research towards the impact of these variable characteristics. Further, as ambient images contain the variable elements inherent in natural environments, findings derived from ambient (relative to standardised) images can be more directly generalised to real-life contexts. Nonetheless, standardised images remain useful in improving sensitivity to detect small effects and in manipulating images to determine causal effects (Sutherland, Rhodes, et al., 2017). A strength of the present research is the complementary use of ambient (Studies 1-2) and more controlled images (Studies 3-4), together offering a more comprehensive insight into person perception (Sutherland et al., 2013; Sutherland, Rhodes, et al., 2017; Todorov, 2017). The fact that we obtained very similar findings with two different types of images lends confidence to our findings.

Our performance measure involved correlating a participant's responses to items presented at a given exposure duration with those of participants seeing the same items for unlimited time. In first impressions research this measure has the advantage of not assuming an objectively correct answer for each item (cf. Sprengelmeyer et al., 2017). However, whilst results can then be generalized to other participants, we cannot measure generalization to other stimuli. Studies using mixed-model designs would therefore be a useful next step (Judd, Westfall & Kenny, 2017).

Finally, our use of Fourier phase-scrambled masks constitutes an advance over previous masking techniques used in minimal exposure paradigms. For example, Todorov et al. (2009) used a single mask (mosaic of facial fragments), Bar et al.'s (2006) masks were random black lines on a grey/white background, and Willis and Todorov (2006) did not use masks. Effective masking is essential to interrupt the processing of images at a set time (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005), preventing retinal persistence. The masks used in Studies 2 and 4 were individually created to match the original images on colour composition, brightness, and spatial frequency content.

To conclude, many relationships begin in contexts where facial impressions form an important source of information (e.g., online dating). Research using the minimal exposure paradigm has not systematically examined traits relevant to partner preferences and has not used the highly variable images encountered in everyday life. By doing both of these, the present research offered a novel and naturalistic approach to examine the salience of trustworthiness, status, and attractiveness in first impressions of ambient faces and youthful-looking averaged faces. Our findings revealed that a single glance of 100ms is sufficient to form a reliable, consensual first impression and that additional time (500ms presentation) did not result in better correspondence with an independent set of time-unconstrained judgements. Even at 33ms presentation, performance was not severely impaired. The

pervasiveness of online images and internet-based romantic (and professional) relationships -

in which individuals may approach another based on a rapid glance at a profile image - make

these findings timely and relevant to contexts within and beyond the romantic domain.

References

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course

of visual processing: Backward masking and natural scene categorisation. *Vision*

Research, *45*(11), 1459–1469. http://doi.org/10.1016/j.visres.2005.01.004

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and

unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*(46),

17948–53. http://doi.org/10.1073/pnas.0705435104

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278.

http://doi.org/10.1037/1528-3542.6.2.269

Baseler, H. A., Harris, R. J., Young, A. W., & Andrews, T. J. (2014). Neural responses to

expression and gaze in the posterior superior temporal sulcus interact with facial

identity. *Cerebral Cortex*, *24*(3), 737–744. http://doi.org/10.1093/cercor/bhs360

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of

familiar faces. *British Journal of Psychology*, *102*(4), 943–958.

http://doi.org/10.1111/j.2044-8295.2011.02039.x

Campbell, L., Simpson, J. A., Kashy, D. A., & Fletcher, G. J. O. (2001). Ideal standards, the

self, and flexibility of ideals in close relationships. *Personality and Social Psychology*

Bulletin, *27*(4), 447–462. http://doi.org/10.1177/0146167201274006

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In *The Adapted*

Mind: Evolutionary Psychology and the Generation of Culture (pp. 163–228).

http://doi.org/10.1098/rstb.2006.1991

Fiore, A. T., Taylor, L. S., Mendelsohn, G. a, & Hearst, M. (2008). Assessing attractiveness

in online dating profiles. *Proceedings of the Twenty-Sixth Annual CHI Conference on*

Human Factors in Computing Systems CHI 08, (1992), 797–806.

http://doi.org/10.1145/1357054.1357181

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in

samples from an indefinitely large population. *Biometrika*, *10*(4), 507–521.

http://doi.org/10.2307/2331838

Fletcher, G. J. O., Kerr, P. S. G., Li, N. P., & Valentine, K. A. (2014). Predicting romantic

interest and decisions in the very early stages of mate selection: Standards, accuracy,

and sex differences. *Personality and Social Psychology Bulletin*, *40*, 540–50.

http://doi.org/10.1177/0146167213519481

Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000a). Ideals, perceptions, and evaluations

in early relationship development. *Journal of Personality and Social Psychology*, *79*(6),

933–940. http://doi.org/10.1037//0022-3514.79.6.933

Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000b). The measurement of perceived

relationship quality components: A confirmatory factor analytic approach. *Personality

and Social Psychology Bulletin*, *26*(3), 340–354.

http://doi.org/10.1177/0146167200265007

Fletcher, G. J. O., Simpson, J. A., Thomas, G., & Giles, L. (1999). Ideals in intimate

relationships. *Journal of Personality and Social Psychology*, *76*(1), 72–89.

http://doi.org/10.1037/0022-3514.76.1.72

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.

http://doi.org/10.1016/j.cub.2015.08.048

Hall, C. C., Goren, A., Chaiken, S., & Todorov, A. (2009). Shallow cues with deep effects:

Trait judgments from faces and voting decisions. In E. Borgida, C. M. Federico, & J. L.

Sullivan (Eds.), *The Political Psychology of Democratic Citizenship* (Vol. 1, pp. 583–

605). New York, NY: Oxford University Press. http://doi.org/10.1093/acprof

Hancock, J. T., & Toma, C. L. (2009). Putting your best face forward: The accuracy of online

dating photographs. *Journal of Communication*, *59*(2), 367–386.

http://doi.org/10.1111/j.1460-2466.2009.01420.x

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. http://doi.org/10.1037/0096-1523.32.2.199

Jasienska, G., Lipson, S. F., Ellison, P. T., Thune, I., & Ziomkiewicz, A. (2006). Symmetrical

women have higher potential fertility. *Evolution and Human Behavior*, *27*(5), 390–400.

http://doi.org/10.1016/j.evolhumbehav.2006.01.001

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of

the same face. *Cognition*, *121*(3), 313–323.

http://doi.org/10.1016/j.cognition.2011.08.001

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random

factor: designs, analytic models, and statistical power. *Annual Review of Psychology, 68*,

601-625. http://doi:10.1146/annurev-psych-122414-033702

Lassek, W. D., & Gaulin, S. J. C. (2008). Waist-hip ratio and cognitive ability: Is

gluteofemoral fat a privileged store of neurodevelopmental resources? *Evolution and

Human Behavior*, *29*(1), 26–34. http://doi.org/10.1016/j.evolhumbehav.2007.07.005

Meetic Group. (2016). Meetic group: Who are we? Retrieved April 19, 2016, from

https://www.meetic-group.com/en/meetic-group-who-are-we/

Møller, A. P. (1999). Developmental stability is related to fitness. *The American Naturalist*,

153(5), 556–560. http://doi.org/10.1086/303197

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory*. *McGraw-Hill Higher Education*

(Vol. 3). http://doi.org/10.1037/018882

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait

inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110.

http://doi.org/10.1007/s10919-009-0082-1

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings

of the National Academy of Sciences of the United States of America*, *105*(32), 11087–

92. http://doi.org/10.1073/pnas.0805664105

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of

Psychology*, *57*, 199–226. http://doi.org/10.1146/annurev.psych.57.102904.190208

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.

http://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. http://doi.org/10.1068/p5278

Santos, I. M., & Young, A. W. (2011). Inferring social attributes from different face regions:

Evidence for holistic processing. *Quarterly Journal of Experimental Psychology*, *64*(4),

751–66. http://doi.org/10.1080/17470218.2010.519779

Sprengelmeyer, R., Young, A. W, Baldas, E-M., Ratheiser, I., Sutherland, C. A. M., Müller,

H-P., et al. (2016). The neuropsychology of first impressions: evidence from

Huntington's disease. *Cortex, 85*, 100-115. https://doi.org/10.1016/j.cortex.2016.10.006

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., &

Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-

dimensional model. *Cognition*, *127*(1), 105–18.

http://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. http://doi.org/10.1016/j.cognition.2016.09.006

Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial image manipulation: A

tool for investigating social perception. *Social Psychological and Personality Science*,

1(14), 194855061769717. http://doi.org/10.1177/1948550617697176

Sutherland, C. A. M., Young, A. W., & Rhodes, G. (2017). Facial first impressions from

another angle: How social judgements are influenced by changeable and invariant facial

properties. *British Journal of Psychology*, *108*(2), 397–415.

http://doi.org/10.1111/bjop.12206

Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Prototyping and transforming facial

textures for perception research. *IEEE Computer Graphics and Applications*, *21*(4), 42–

50. http://doi.org/10.1109/38.946630

Todd, P. M., Penke, L., Fasolo, B., & Lenton, A. P. (2007). Different cognitive processes

underlie human mate choices and mate preferences. *Proceedings of the National*

Academy of Sciences of the United States of America, *104*(38), 15011–15016.

http://doi.org/10.1073/pnas.0705290104

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton, NJ:

Princeton University Press.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence

from faces predict election outcomes. *Science*, *308*(5728), 1623–1626.

http://doi.org/10.1126/science.1110589

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*, 519–545. http://doi.org/10.1146/annurev-psych-113011-

143831

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness

after minimal time exposure. *Social Cognition*, *27*(6), 813–833.

http://doi.org/10.1521/soco.2009.27.6.813

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different

facial images of the same person. *Psychological Science*, *25*(7), 1404–1417.

http://doi.org/10.1177/0956797614532474

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first

impressions from highly variable facial images. *Proceedings of the National Academy of*

Sciences of the United States of America, *111*(32), E3353-61.

http://doi.org/10.1073/pnas.1409860111

Walster, E., Aronson, V., Abrahams, D., & Rottman, L. (1966). Importance of physical

attractiveness in dating behavior. *Journal of Personality and Social Psychology*, *4*(5),

508–516. http://doi.org/10.1037/h0021188

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.

http://doi.org/10.1111/j.1467-9280.2006.01750.x