Beauty *is* in the eye of the Beholder: The Appraisal of Facial Attractiveness and its relation to Conscious Awareness

Myron Tsikandilakis[1,2]

Persefoni Bali[1]

Peter Chapman[1]

[1]School of Psychology, University of Nottingham

[2]Medical School, Faculty of Medicine and Health Sciences, University of Nottingham

Address for correspondence:

Tsikandilakis Myron

School of Psychology, University of Nottingham, lpxmt2@nottingham.ac.uk, myron.tsikandilakis@kcl.ac.uk , ps202mt@gold.ac.uk .

Key words: attractiveness, masking, awareness

Word Count: 7.733

Abstract

Previous research suggests that facial attractiveness relies on features such as symmetry, averageness and above-average sexual dimorphic characteristics. Due to the evolutionary and sociobiological value of these characteristics, it has been suggested that attractiveness can be processed in the absence of conscious awareness. This raises the possibility that attractiveness can also be appraised without conscious awareness. In the present study, we addressed this hypothesis. We presented neutral and emotional faces that were rated high, medium and low for attractiveness during a pilot experimental stage. We presented these faces for 33.33 ms with backwards masking to a black and white pattern for 116.67 ms and measured face-detection and emotion-discrimination performance, and attractiveness ratings. We found that high-attractiveness faces were detected and discriminated more accurately and rated higher for attractiveness compared to other appearance types. A Bayesian analysis of signal detection performance indicated that faces were not processed significantly at-chance. Further assessment revealed that correct detection (hits) of a presented face was a necessary condition for reporting higher ratings for high-attractiveness faces. These findings suggest that the appraisal of attractiveness requires conscious awareness.

Introduction

Attractiveness is considered a fundamental aspect of human interactions from infancy (Fisher & Ma, 2014; Thomas et al., 2007) to adulthood (Griffin & Langlois, 2006). It has been suggested to positively influence the quality of our peer relations, our behaviour and the behaviour of others towards us throughout our life (Langlois & Stephan, 1981), to influence and inform our mate choices (Saxton, Caryl & Roberts, 2006), and even exert influence in our professional development and income status (Frieze, Olson & Russell, 1991). In our modern society, where interpersonal communication, romantic-mate choice and even professional employability rely increasingly on on-line communication and posted photographic material (Hong, Tandoc Jr, Kim E., Kim B., & Wise, 2012), the perception of facial attractiveness has been re-approached as a highly relevant psychological research subject (Swami & Farnham, 2008).

Facial attractiveness has been associated with certain perceived characteristics; these include – among others (Little, Jones & DeBruine, 2011) – symmetry (Scheib, Gangestas & Thornhill, 1999), averageness (Grammer & Thornhill, 1994) and above-average sexually dimorphic characteristics (Barber, 1995). Bilateral facial symmetry, for example, is assumed to function as a cue for developmental health and ontogenetic resistance to parasites (Fink, Neave, Manning & Grammer, 2006). Facial averageness is also suggested to confer evolutionary important offspring-survival value due to indications of heterozygosity (Roberts et al., 2005) as well as a developmental propensity for familiarity due to average faces being overall more recognisable and prototypical exemplars of the *category* facial stimuli (Fink & Penton-Voak, 2002). Additionally, pronounced facial characteristics that signify sexually dimorphic biological markers, such as testosterone in men and oestrogens in women, have been shown to increase gender-specific attractiveness ratings (Smith et al., 2006).

Although recent research has reported evidence that casts doubt on the role of attractiveness particularly as an indication of developmental health and immunocompetence (Scott, Clark, Boothroyd, & Penton-Voak, 2013; Foo, Simmons & Rhodes, 2017; Jones et al., 2017; Cai et al., 2018), the association of attractiveness with characteristics that could confer evolutionary important sociobiological value was sufficient to prompt researchers to explore whether attractiveness can be processed, reported and appraised under conditions that do not necessarily involve conscious awareness (Mineka & Öhman, 2002; Öhman, 2009). This line of research has suggested that attractiveness is a highly-salient visual characteristic (Barber, 1995) that can be appraised even from impoverished (pixelated) visual cues (Bachmann, 2007). Previous research has also suggested that – due to the evolutionary importance of high-attractiveness features – attractiveness can be processed and reported at a glance from faces, after only 13 ms of visual exposure to a high-attractiveness face (Olson & Marshuetz, 2005), and has reported evidence for inhibition of return to lateral Gabor patches preceded by high-attractiveness faces that were adjusted for visibility using staircase visual contrast manipulations (Hung, Nieh & Hsieh, 2016). Based on these findings, the argument has been raised that attractiveness "can be processed in the absence of conscious awareness" (Hung et al., 2016; p. 6) and that it can be appraised and reported without conscious awareness (Olson & Marshuetz, 2005).

In the current study, we explored whether particularly the appraisal of attractiveness requires conscious awareness. We presented emotional and neutral faces that were rated high, medium and low in attractiveness. We presented these faces for 33.33 ms (Brooks et al., 2012) using backwards masking to a black and white pattern for 116.67 ms. We measured face-detection and emotion-discrimination performance as well as attractiveness ratings in response to these faces. To accurately assess these responses, we used the paradigm we have developed for the assessment of subliminality described in detail in previous research (Tsikandilakis, Chapman & Peirce, 2017; Tsikandilakis & Chapman, 2018). This included response assessment using unbiased non-parametric signal detection theory criteria to measure face-detection and emotion-discrimination performance (Stanislaw & Todorov, 1999; Zhang & Mueller, 2005), Bayesian analysis (Dienes, 2015) for the assessment of chance-level significance (A = .5) that would indicate stimuli invisibility (Erdelyi, 2004), and hits (correct detection/discrimination) and misses (incorrect detection/discrimination) analysis of participant ratings (Fawcett, 2006).

The inclusion of neutral faces (Olson & Marshuetz, 2005; Hung et al., 2016) as well as emotional faces in the current study was necessitated by signal detection theory requirement for the exploration of different levels of conscious awareness (Stanislaw & Todorov, 1999). The inclusion of different emotions enabled us to explore whether the appraisal of attractiveness requires face detection as well as emotion discrimination of the presented face. This signal detection theory distinction could not have been performed without the inclusion of different emotions (Pessoa, 2005). This inclusion also meant that the current study would be the first to our knowledge study (Olson & Marshuetz, 2005; Hung et al., 2016; Ritchie, Palermo & Rhodes. 2017) that included assessment of different emotional expressions and appearance types (high, medium and low-attractiveness faces) under conditions of backwards masking.

This experimental condition allowed us to formulate a secondary and – in the absence of previous research in the area (Fink & Penton-Voak, 2002) – exploratory hypothesis. We explored whether the interaction between emotional expressions and appearance types can influence face detection, emotion discrimination, and attractiveness ratings under conditions of backwards masking, in the same manner that previous research has reported that it influences participant responses during supraliminal presentations (O'Doherty et al., 2002; Calvo & Lundqvist, 2008; Lundqvist, Bruce & Öhman, 2015). The preliminary hypothesis for this exploratory objective was that faces that have high sociobiological value based on attractiveness ratings, i.e. high-attractiveness faces (Bachmann, 2007), would reveal higher face-detection and emotion-discrimination scores when they expressed emotions that also have high sociobiological and survival-related evolutionary value (Brooks et al., 2012), such as fearful (Pessoa, 2005) and angry faces (Lundqvist, Bruce & Öhman, 2015). We also explored whether higher attractiveness ratings would be reported for high-attractiveness faces that expressed highly-salient positive social signals, such as happy expressions, that have been associated with increased emotional-perceptual reward value in previous research (O'Doherty et al., 2002; Calvo & Lundqvist, 2008).

In the current research, it was extremely important that any potential influences of attractiveness could not be interpreted in terms of other evolutionarily relevant or low-level stimulus features. We thus included three pilot experimental stages (Study One) to ensure that the stimuli we used did not confound attractiveness with gender, emotional expression, or detectability through contrast changes. In the first experimental stage, we validated the facial stimuli for gender and emotional characteristics and used strict attractiveness criteria to pre-select from an existing database (Gur et al., 2002) faces that were rated high, medium and low in attractiveness. In the second experimental stage, we assessed the pre-selected stimuli for emotionality to make sure that emotionality differences between appearance types would not bias signal detection performance in subsequent experimental stages (Calvo & Lundqvist, 2008). Finally, in stage three, we explored whether there were differences in visual contrast between the selected faces and their control condition during the main experiment (non-facial pattern stimuli) that could artefactually impact signal detection and discrimination responses when using backwards masking (Bachmann & Francis, 2013).

Methods

Study One

Stage One: Stimuli Pre-Selection

Aims. The current stage had two aims. The first aim was to select from an existing database (Gur et al., 2002) the faces that were correctly recognised by participants and automatic facial recognition software (Noldus) for the emotion that they were expressing. The second aim was to select faces that were rated high, medium and low in attractiveness and test whether these faces produced significant differences in attractiveness that would make them appropriate stimuli for the inclusion in the following experimental stages.

Participants. A power calculation based on effect sizes (d = .81; f = .41) reported in previous research (Tsikandilakis et al., 2017) revealed that fifteen participants were required for $P_{(1-\beta)} \geq .8$ (Faul, Erdfelder, Buchner & Lang, 2009). Eighteen participants (nine females) volunteered to participate in this experiment. All participants reported normal or corrected-to-normal vision. Participants gave informed consent to participate in the current study prior to the experiment. The participants were screened before the experiment with the Somatic and Psychological Health Report Questionnaire (Hickie et al., 2001); participants with scores at or below 1.0 were included. Participants were also screened using an online Alexithymia-Emotional Blindness questionnaire (Alexithymia, 2017) and participants with scores that indicated possible traits ($P > 94$) or diagnosis ($P > 112$) for alexithymia were excluded; data from a single participant were excluded from the study. Two participants were also excluded from the study due to neutral ratings on the attractiveness task (Olson & Marshuetz, 2005; Hung et al., 2016). The final population sample consisted of fifteen participants (eight females) with mean age 32.87 (S.D. = 6.12). The experiment was approved by the Ethics Committee of the School of Psychology of the University of Nottingham.

Stimuli and Procedures. The facial stimuli used were taken from the dataset created by Gur and colleagues (2002) and included faces with angry, happy, fearful, sad and neutral facial expressions. The stimuli were adjusted for interpupillary distance, transformed to grey scale and resized to a standard 1024x768 pixels resolution. Their luminescence was averaged in SHINE, MATLAB Toolbox and Fourier Painter and finally they were spatially aligned and framed into pure white within a cropped circle (Height: 6 cm, Width: 4 cm).

A total of three-hundred faces were presented from sixty different actors. Ninety cropped non-facial blurs patterns that were matched for luminescence (SHINE, MATLAB Toolbox) with the presented faces were also shown. The experimental trial started with a fixation cross for three seconds (± one second). After the fixation cross in random order a single face or non-facial pattern was presented at fixation for one second. After each target a black and white pattern mask was also presented for one second. A blank screen interval was then presented for two seconds. After that participants were asked by an on-screen message to rate how attractive the presented stimulus was from one (not attractive at all) to ten (very attractive) using the keyboard. Participants were also asked to decide from an on-screen list what kind of stimulus was presented during the trial using the keyboard. The list included (a) angry, (f) fearful, (h) happy, (s) sad, (n) neutral, (o) other and (i) non-facial. The order of the two engagement tasks was randomised in each trial. A two-second blank screen interval was presented before the next trial.

Stimuli pre-selection. We selected from the presented faces the ones that reported 100% accuracy in correct discrimination of emotional expression (n = 248). These stimuli were further analysed using Noldus Face Reader 7.0 to validate their emotion. We used the participant calibration module for emotional recognition that controlled for the action units that were present in the neutral expressions of each actor to accurately assess emotional expressions. We also used the cultural-background recognition module and specific cultural-background emotional recognition modules (e.g. General6[1], Asian etc.) for each actor based on the cultural-background recognition assessment (Noldus, 2018). We set the emotional recognition certainty criteria for inclusion at > .99 for each facial stimulus; no stimuli were excluded (Appendix 1.1). From the resulting dataset we chose for each emotion (anger, fear, happiness, sadness and neutral) faces that were rated high, medium and low in attractiveness based on the following criteria. Six (three male and three female) high, medium and low-attractiveness faces were selected for each emotion. For high-attractiveness faces, the faces with a mean value that was more than seven were pre-selected. For medium-attractiveness faces we pre-selected faces that were rated between four and six on the attractiveness scale. For low-attractiveness faces we pre-selected faces that were not rated higher than three on the attractiveness scale. Due to rating restrictions and to avoid identity priming due to uneven target repetition (Lander, Bruce & Hill, 2001), in subsequent stages actors that met the required attractiveness criteria per emotion were selected three times each, resulting in a final sample of thirty actors and ninety emotional expressions.

Results and Discussion: To confirm that attractiveness was different between different appearance types we ran a repeated measures ANOVA with independent variables Appearance Type (high, medium and low attractiveness) and Type of Emotion (anger, fear, happiness, sadness, neutral) and dependent variable attractiveness ratings. Appearance Type was significant (F (2, 28) = 1739.43, $p < .01$, $\eta^2 = .99$) confirming that high-attractiveness faces (M. = 7.8, S.D. = .23) were rated higher (t (14) = 30.01, $p < .01$; $d = 9.24$) than medium-attractiveness faces (M. = 5.17, S.D. = .33) and higher (t (14) = 74.08, $p < .01$; $d = 26.55$) than low-attractiveness faces (M. = 2.54, S.D. = .16) in attractiveness ratings. Medium-attractiveness faces were also rated higher than low-attractiveness faces (t (14) = 46.29, $p < .01$; $d = 10.14$) in attractiveness ratings. The findings suggested that appearance types (high, medium and low) were significantly different in attractiveness ratings and that they were appropriate stimuli for their inclusion in the following experimental stages.

In accordance with previous literature in the area (Olson & Marshuetz, 2005; Hung et al., 2016) we did not find any differences for actor gender attractiveness ratings (F (1, 13) = .11; p = .75; η^2 = .01) or with participant gender as a between-subjects variable (F (1, 13) = .37; p = .56; η^2 = .03). These findings were confirmed by a separate t-test analysis based on actor (t (14) = - .473; p = .64; d = .13) and participant gender attractiveness ratings (t (13) = .407; p = .69; d = .18). These findings suggested that male and female participants did not differ in their ratings for attractiveness (see also Appendix 2.1 and 3.1).

Figure 1: Attractiveness Ratings per Appearance Type and Expressed Emotion

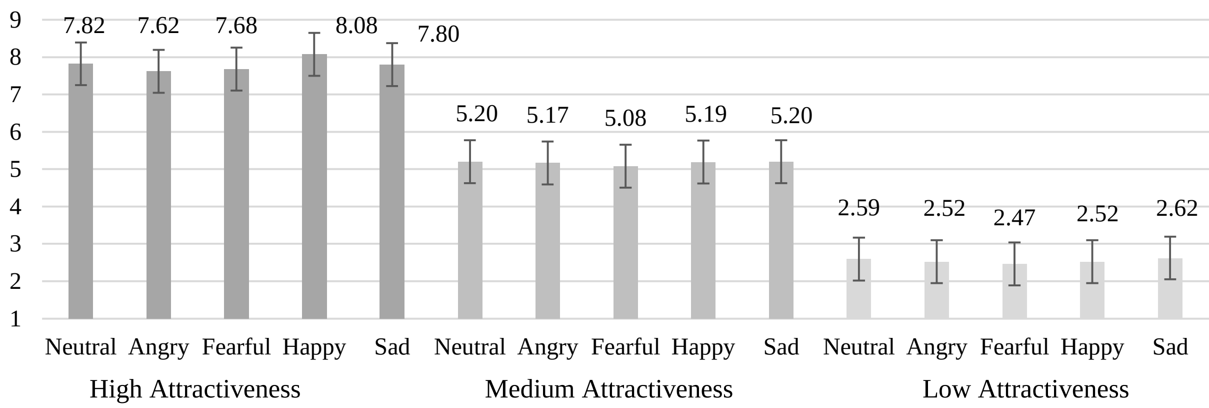


Figure 1: Mean attractiveness ratings (Y axis) per appearance type and expressed emotion (X Axis). Error bars indicate standard error of the mean (±2).

Stage Two: Emotionality Assessment

Aims. The aim of this stage was to assess the faces that were rated high, medium and low in attractiveness for emotionality differences to make sure that emotionality differences between appearance types would not bias signal detection performance in subsequent experimental stages.

Participants. A power calculation based on effect sizes (d = .86; f = .43) reported in previous research (Tsikandilakis et al., 2017) revealed that thirteen participants were required for $P_{(1-\beta)}$

≥ .8. Fifteen participants (seven females) that were not part of stage one volunteered to participate in this stage. All participants reported normal or corrected-to-normal vision and gave informed consent to participate in the current study prior to the experiment. The participants were screened with the Somatic and Psychological Health Report Questionnaire and an online Alexithymia-Emotional Blindness questionnaire; no participants were excluded. Data from two participants were excluded due to neutral ratings on the emotionality task (Olson & Marshuetz, 2005; Hung et al., 2016). The final population sample consisted of thirteen participants (eight females) with mean age 24.85 (S.D. = 3.95). This stage was approved by the Ethics Committee of the School of Psychology of the University of Nottingham.

Stimuli and Procedures. A total of ninety faces were shown during this stage from thirty actors. An equal number of female and male faces (n = 45) and actors (n = 15) were presented. Six faces (three male and three female) were shown per emotion (anger, fear, happiness, sadness and neutral) for each appearance type (high, medium and low attractiveness). The faces presented during this stage were the pre-selected stimuli from stage one. The ninety non-facial pattern blurs that were shown during stage one were also shown during this stage. The experimental trial started with a fixation cross for three seconds (± one second). After the fixation cross in random order a single face or non-facial pattern was presented at fixation for one second. After each target a black and white pattern mask was also presented for one second. A blank screen interval was then presented for two seconds. Participants were then asked by an on-screen message to rate how emotional the presented stimulus was from one (not emotional at all) to ten (very emotional) using the keyboard. A two-second blank screen interval was presented before the next trial.

Results and Discussion. A repeated measures analysis of variance tested the effects of Appearance Type (high, medium and low attractiveness) and Type of Emotion (angry, fearful, happy, sad and neutral) on emotional ratings. The analysis revealed that there were no significant differences in emotional ratings between high, medium and low-attractiveness faces (F (2, 24) = 2.94, p = .12; η^2 = .16; see also Appendix 4.1). These findings suggested that high, medium and low-attractiveness faces were not overall different in emotional ratings and were appropriate stimuli for their inclusion in subsequent experimental stages.

An analysis of variance revealed that there were no differences for actor gender in emotionality ratings (F (1, 11) = .64; p = .44; η^2 = .06) or with participant gender as a between-subjects variable (F (1, 11) = .1; p = .75; η^2 = .01). This was confirmed by a separate t-test analysis based on actor (t (12) = - .77; p = .12; d = .05) and participant gender emotionality ratings (t (11) = 1.41; p = .19; d = .73). These findings suggested that male and female participants did not differ in their ratings for emotionality (see also Appendix 5.1).

Figure 2: Emotionality Ratings per Appearance Type and Expressed Emotion

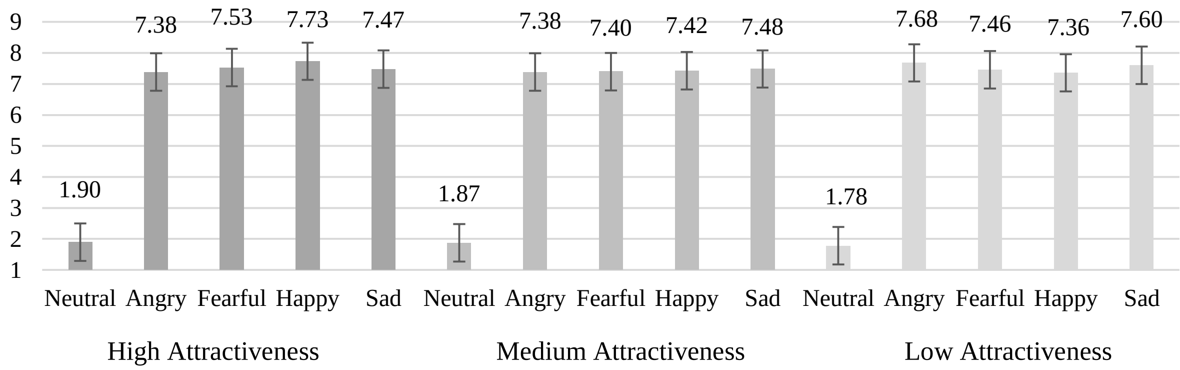


Figure 2: Mean emotionality ratings (Y axis) per appearance type and expressed emotion (X Axis). Error bars indicate standard error of the mean (±2).

Stage Three: Subjective Contrast Discrimination

Aims. Study two was designed to include a 116.67 ms black and white pattern mask and 33.33 ms non-facial blurs as a signal detection control condition for facial stimuli. For such a control condition to work it is important that low level changes in contrast cannot be used by participants to detect the presence of faces. The aim of the current stage was to ensure that brief non-facial blurs did not have differences in visual contrast in respect to the black and white pattern mask when compared to the facial stimuli.

Participants: A power calculation based on effect sizes[1] (d = .05; f = .03) reported in previous research (Tsikandilakis et al., 2017) revealed that fourteen participants were required for $\beta < .2$ (Wesa, 2016; Soper, 2018). Fifteen participants (seven female) that were not part of experiment one or two volunteered to participate in this experimental stage. All participants reported normal or corrected-to-normal vision and provided informed consent prior to the study. The participants were screened with the Somatic and Psychological Health Report Questionnaire and an online Alexithymia-Emotional Blindness questionnaire; no participants were excluded based on this assessment. One participant was excluded due to non-compliance with the study procedures. The final population sample consisted of fourteen participants (seven female) with a mean age of 23.07 (S.D. = 2.70). Participants were briefed in writing concerning the experimental task and were asked to respond in the consent form whether they understood the instructions (yes or no). All participants responded positively. This stage was approved by the Ethics Committee of the School of Psychology of the University of Nottingham.

Stimuli and Procedures: Due to experimental time restrictions a subset of the selected faces including thirty faces and thirty non-facial blurs were presented during this stage. The presented faces were randomly chosen from the pre-selected stimuli in stages one and two. The selected set was assigned two faces per emotion (neutral, angry, fearful, happy and sad) for each appearance type (high, medium and low attractiveness). An equal number of male and female faces (n = 15) were presented to participants and no actor was repeated more than once at this stage. The experimental trial started with a fixation cross for three seconds (± one second). After the fixation cross in random order a single face or non-facial blur was presented for 33.33 ms followed by a black and white patterned mask for 116.67 ms. After the presentation of the black and white pattern mask, participants were shown a blank screen for two seconds and were

[1] See Schmider, Ziegler, Danay, Beyer and Bühner (2010)

then asked by an on-screen message to rate their subjective experience of visual contrast from one (not at all) to ten (intense). A two-second blank screen interval was presented before the next trial.

Results and Discussion. To test if the non-facial blurs had significantly different ratings for subjective experience of contrast compared to the presented faces in respect to the pattern mask a paired samples t-test was ran. Subjective experience of contrast for the non-facial blurs (M. = 4.95, S.D. = .49) was not rated higher than contrast in the face condition (M. = 4.96, S.D. = .21; t (13) = -.119, p = .907; d = .03; see also Appendix 6.1). These findings suggested that differences of visual contrast between the non-facial blurs and the presented faces, and the pattern mask would not artefactually impact signal detection and discrimination performance (Bachmann & Francis, 2013) in subsequent experimental stages.

Study Two

Aims. The primary aims of this study were twofold. Firstly, we wanted to test whether high-attractiveness faces would be detected and discriminated more accurately than other appearance types under conditions of backwards masking. Secondly, we wanted to test whether the appraisal of attractiveness in high-attractiveness faces requires conscious awareness. Finally an exploratory aim of the current study was to test if high, medium and low-attractiveness faces interact with different types of emotional expressions (fearful, angry, happy, sad and neutral) to influence face-detection and emotion-discrimination performance as well attractiveness ratings under conditions of backwards masking.

Participants. A power calculation based on medium effect sizes ($\eta^2 = .06$; f = .25) revealed that twenty-three participants would be required for $P_{(1-\beta)} \geq .8$. Twenty-six participants (thirteen females) that were not part of Study One volunteered to participate in this experiment. All participants reported normal or corrected-to-normal vision and gave informed consent to participate in the current study prior to the experiment. The participants were screened with the Somatic and Psychological Health Report Questionnaire (Hickie et al., 2001) and an online Alexithymia-Emotional Blindness questionnaire (Alexithymia, 2017); data from one participant were excluded due to possible Alexithymia traits (> 94). Data from two additional participants were also excluded due to neutral ratings on the pre-experimental emotionality task and attractiveness tasks (Olson & Marshuetz, 2005; Hung et al., 2016; van der Ploeg et al., 2017). The final population sample consisted of twenty-three participants (thirteen females) with mean age 32.13 (S.D. = 7.47). This experiment was approved by the Ethics Committee of the School of Psychology of the University of Nottingham.

Stimuli and Procedures. The experiment involved two phases. In phase one participants were presented with twenty faces and twenty non-facial pattern blurs that were not part of the pre-selected stimuli. These stimuli were chosen based on discrimination, emotionality, intensity, expression-ambiguity ratings, and physiological responses (skin conductance and heart rate) in a previous study (Tsikandilakis et al., 2017). These facial stimuli included four neutral, angry, happy, sad and fearful stimuli from different actors with an equal number (n = 10) of males and females. The experimental trial started with a fixation cross for three seconds (± one second). After the fixation cross in random order a single face or non-facial pattern was presented at fixation for one second. After each target a black and white pattern mask was also presented for one second. A blank screen interval was then presented for two seconds. After that participants were asked by an on-screen message to rate how emotional the presented stimulus was from one (not emotional at all) to ten (very emotional) using the keyboard. The participants were also asked by an on-screen message to rate how attractive the presented stimulus was from one (not attractive at all) to ten (very attractive) using the keyboard. The participants were also asked to decide from an on-screen list what kind of stimulus was presented during the trial using the keyboard. The list included (a) angry, (f) fearful, (h) happy, (s) sad, (n) neutral, (o) other and (i) non-facial. The order of the engagement tasks was randomised in each trial. A blank screen for two seconds was presented before each next trial. Phase one was conducted to ensure that participants were familiar with the tasks and stimuli under conditions in which the target stimuli were clearly visible for all participants. In phase two we used the same procedure with brief backwardly masked stimuli presented at durations where they would not necessarily be available to conscious awareness (Dehaene et al., 2006).

After phase one participants were allowed a five-minute break. After the break the participants were presented with the ninety pre-selected faces and ninety non-facial pattern blurs. The experimental trial started with a fixation cross for three seconds (± one second). After the fixation cross in random order a single face or non-facial pattern was presented at fixation for 33.33 ms. After each target a black and white pattern mask was also presented for 116.67 ms. A blank screen interval was then presented for two seconds. After that participants were asked to reply to a set of engagement tasks with order randomised. They were asked by an on-screen message to press E if they saw a facial stimuli or W if the presented target was non-facial; the assignment of the keyboard responses was randomly counterbalanced in each trial. After this initial task we used conditional branching to present the participants with additional engagement tasks. If the participant responded having seen a facial stimulus, an on-screen message asked participants to decide from a list what kind of emotion the facial stimulus was expressing using the keyboard. The list included (a) angry, (f) fearful, (h) happy, (s) sad, (n) neutral and (o) other. If the participants replied not having seen facial stimulus, an on-screen message asked them to decide what kind of emotion best described their experience during the presentation using the keyboard. The list included (a) anger, (f) fear, (h) happiness, (s) sadness, (n) neutral and (o) other. Participants were also asked by an on-screen message to rate how attractive the presentation was from one (not attractive at all) to ten (very attractive) using the keyboard (Figure 3). A blank screen for two seconds was presented before the next trial.

Apparatus and Presentation Testing. All experiments were generated using the coder and builder components of Psychopy version 1.90d (Peirce, 2007). All stimuli for all experimental stages were presented on a standard 60 Hz Toshiba monitor in the same quiet laboratory space. To ensure that particularly brief stimuli (33.33 and 116.67 ms) were correctly presented, an IPAD PRO camera with 120 Hz refresh rate (8.33 ms) recorded two pilot runs for Study One (Stage Three) and Study Two. The stimuli presentation was assessed frame by frame; no instances of dropped frames were detected. Subsequently, a self-developed dropped frame report script with one frame (16.67 ms) tolerance threshold was coded in Python and two pilot experimental diagnostic sessions were run. The presenting monitor reported no dropped frames; prognostic dropped frame rate was estimated at 1/5000 trials. Experimental stages were, subsequently, run using dropped frames diagnostics; no instances of dropped frames were reported.

Figure 3: Example Stimuli Sequence with High-Attractiveness Happy Masked Face

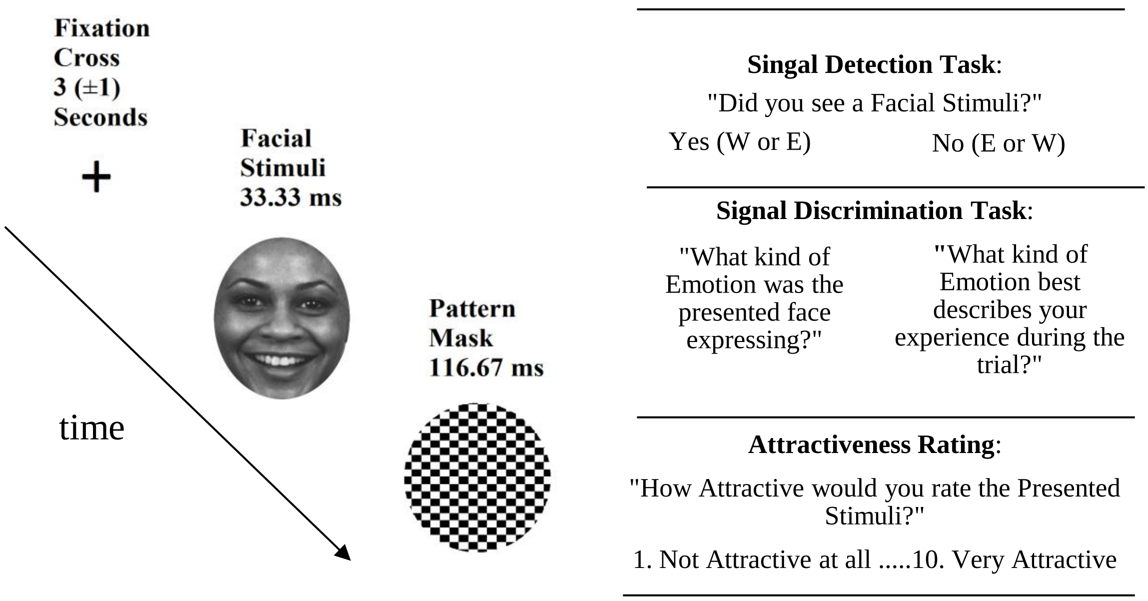


Figure 3: During the main experimental stage participants were presented with thirty high-attractiveness, thirty medium-attractiveness and thirty low-attractiveness emotional faces (angry, fearful, happy, sad and neutral) and ninety pattern blurs for 33.33 ms. Subsequently, they were asked to make face-detection and emotion-discrimination, and attractiveness-rating responses.

Results and Discussion: Does Attractiveness influence Face Detection under conditions of Backwards of Masking? To test whether high-attractiveness faces are detected and discriminated more accurately than other appearance types, the participants' face detection responses were transformed to non-parametric sensitivity index A (Zhang & Mueller, 2005). An analysis of variance with independent variables Appearance Type (high, medium and low attractiveness) and Type of Emotion (angry, fearful, happy, sad, neutral) and dependent variable face-detection performance (A) was ran. The analysis revealed a significant effect of Appearance Type ($F (2, 44) = 9.49$, $p < .01$; $\eta^2 = .3$) and a significant effect of Type of Emotion ($F (2.81, 61.79) = 5.27$, $p < .01$; $\eta^2 = .19$; Greenhouse-Geisser corrected). Further Bonferonni corrected pairwise comparisons revealed that high-attractiveness faces (M. (A) = .797, S.D. (A) = .088; M. (H.R.[2]) = 74.07 % S.D. (H.R.) = 14.64 %) were detected more accurately than medium-attractiveness faces (M. (A) = .768, S.D. (A) = .093; M. (H.R.) = 69.72 % S.D. (H.R.) = 20.77 %; $p < .01$; $d = .32$.). We did not observe any gender effects ($F (2, 42) = .71$, $p = .49$; $\eta^2 = .03$).

Emotion discrimination was also calculated using non-parametric sensitivity index A (Zhang & Mueller, 2005). An analysis of variance was run to assess the effects of Appearance Type and Type of Emotion with dependent variable emotion-discrimination performance (A). The analysis revealed a significant effect of Appearance Type ($F (2, 44) = 4.97$, $p = .01$; $\eta^2 = .18$), a significant effect of Type of Emotion ($F (40.6, 2.57) = 12.58$, $p < .01$; $\eta^2 = .36$), and a significant interaction ($F (8, 176) = 10.81$, $p < .01$; $\eta^2 = .32$). Further Bonferonni corrected pairwise comparisons revealed that high-attractiveness faces (M. (A) = .756, S.D. (A) = .082; M. (H.R.) = 66.82 % S.D. (H.R.) = 20.37 %) were discriminated better than low-attractiveness faces (M. (A) = .734, S.D. (A) = .087; M. (H.R.) = 63.78 % S.D. (H.R.) = 20.51 %; $p < .01$; d

[2] H.R.: Hit Rates

= .26). We did not observe any gender effects (F (2, 42) = 1.65, p = .2; η^2 = .07). These results suggested that attractiveness influenced face detection and emotion discrimination under conditions of backwards masking and more specifically that high-attractiveness faces were detected more accurately than medium-attractiveness faces and discriminated more acutely than low attractiveness-faces.

Exploratory Analysis. The exploratory analysis in the current stage tested whether appearance type and emotion interact under conditions of backwards masking to influence face-detection and emotion-discrimination performance. The analysis revealed a significant interaction between Appearance Type and Emotion (F (8, 176) = 10.81, p < .01; η^2 = .32). Further Bonferonni corrected pairwise comparisons revealed that fearful (p < .01; d = .36) and sad (p < .01; d = 42) high-attractiveness faces were detected more accurately than other appearance type to emotion combinations, and that fearful (p < .01; d = 36) and angry (p < .001; d = .65) high-attractiveness faces, and also neutral low-attractiveness faces (p < .001; d = .99) were discriminated more accurately than other appearance type to emotion combinations. These results suggested that a significant interaction between attractiveness and emotion influenced face detection and emotion discrimination for the presented faces (Table 1).

Table 1: Detection and Discrimination Performance per Appearance Type and Emotion

Appearance Type	Type of Emotion	Mean Detection (A)	S.D. Detection (A)	Mean Discrim. (A)	S.D. Discrim. (A)	Detection Standard. Cohen's d	Discrim. Standard. Cohen's d
High-Attractiveness	Fearful	.818	.132	.782	.101	.357*	.355*
	Angry	.798	.094	.815	.102	.161	.654**
	Happy	.752	.127	.687	.102	-.281	-.535
	Sad	.825	.094	.747	.115	.416*	.026
	Neutral	.79	.08	.75	.09	.085	.05
Medium-Attractiveness	Fearful	.756	.112	.725	.11	-.242	-.184
	Angry	.788	.095	.76	.102	.067	.14
	Happy	.751	.113	.701	.098	-.287	-.409
	Sad	.777	.139	.771	.11	-.044	.245
	Neutral	.766	.093	.716	.099	-.151	-.266
Low-Attractiveness	Fearful	.77	.098	.717	.128	-.108	-.256
	Angry	.809	.094	.734	.125	.267	-.099
	Happy	.742	.102	.683	.102	-.378	-.576
	Sad	.809	.1	.729	.139	.262	-.143
	Neutral	.768	.090	.851	.086	-.125	.996**

Table 1: Mean and S.D. (Standard Deviation) Detection and Discrimination Performance (A) per Appearance Type and Type of Emotion. In the two end-right columns the standardised effect size (Cohen's d) per stimuli type in units of standard deviations from the overall mean for detection (M. = .781, S.D. = .104) and discrimination performance (M. = .745, S.D. = .107). Asterisk (*) indicates significance at < .01 level. Double asterisk (**) indicates significance at < .001 level.

Results and Discussion: Does the appraisal of Attractiveness require conscious awareness?

To test if high-attractiveness faces can be processed without conscious awareness, a Bayesian analysis (Dienes, 2015) with corrected degrees of freedom (df < 30; SE = (SE x ($\left(1+\frac{20}{dfxdf}\right)$)) (Berry, 1996)) and higher and lower bounds set at .6 and .4 with .5 indicating absolute chance-level performance was run. The analysis revealed that face detection (A) for high-attractiveness (M. = .796, S.E. = .019; B > 3), medium-attractiveness (M. = .768, S.E. = .017; B > 3) and low-attractiveness (M = .779, S.E. = .02; B > 3) faces, and emotion-discrimination performance for high-attractiveness (M = .756, S.E. = .018; B > 3), medium-attractiveness (M = .743, S.E. = .021; B > 3) and low-attractiveness (M =. 734, S.E. = .019; B > 3) faces were above chance-level (Dienes, 2015) that would indicate stimuli invisibility (Erdelyi, 2004). An analysis of overall attractiveness ratings was not appropriate for the exploration of whether the appraisal of attractiveness requires conscious awareness (Figure 5a and 5b).

To further explore whether attractiveness could influence ratings without conscious awareness, we ran an analysis of hits and misses per appearance type for face-detection and emotion-discrimination performance (Pessoa, 2005). For face-detection performance a factorial ANOVA with independent variables Appearance Type (high, medium and low attractiveness), Emotional Type (fearful, angry, happy, sad and neutral) and Detection Response (hits, misses) and dependent variable attractiveness ratings was performed. The analysis revealed a significant effect of Appearance Type ($F(2, 22) = 122.14$, $p < .01$; $\eta^2 = .92$), a significant effect of Emotion ($F(4, 44) = 3.37$, $p = .02$; $\eta^2 = .23$) and a significant effect of Detection Response ($F(1, 11) = 51.47$, $p < .01$; $\eta^2 = .63$). Critically, an Appearance Type by Detection Response interaction was reported ($F(2, 22) = 167.08$, $p < .01$; $\eta^2 = .94$), suggesting that there were attractiveness-rating differences between hits and misses for different appearance types.

To further explore these findings Bonferonni corrected pairwise comparisons were run. The comparisons revealed that for face detection high-attractiveness facial-hits (M. = 7.21, S.D. = .42) were rated higher than medium-attractiveness facial-hits (M. = 4.79, S.D. = .24; $t(22) = 30.55$, $p < .01$; $d = 7.07$) and low-attractiveness facial-hits (M. = 3.78, S.D. = .54; $t(22) = 22.29$, $p < .01$; $d = 7.09$). No significant differences in attractiveness ratings were reported for misses for face detection ($F(1.08, 20.5) = 1.10$, $p = .38$; $\eta^2 = .06$; Greenhouse-Geisser corrected). For detection-misses, high-attractiveness faces were not significantly different than medium-attractiveness ($p = 1$; $d = .22$) and low-attractiveness faces ($p = .96$; $d = .19$) and medium-attractiveness faces were not significantly different than low-attractiveness faces ($p = .85$; $d = .1$).

A similar pattern was reported for emotion-discrimination responses. An analysis of variance was run with independent variables Appearance Type (high, medium and low attractiveness), Emotion (fearful, angry, happy, sad and neutral) and Discrimination Response (hits, misses) with attractiveness ratings as the dependent variable. The analysis revealed a significant effect of Appearance Type ($F\ (2, 14) = 74.01$, $p < .01$; $\eta^2 = .91$), a significant effect of Emotion ($F\ (4, 28) = 2.98$, $p = .04$; $\eta^2 = .29$) and a significant effect of Discrimination Response ($F\ (1, 7) = 132.55$, $p < .01$; $\eta^2 = .95$). Critically, a significant Appearance Type by Discrimination Response interaction was revealed ($F\ (2, 14) = 39.78$, $p < .01$; $\eta^2 = .85$).

To further explore these findings Bonferonni corrected pairwise comparisons were run. For emotion-discrimination performance, high-attractiveness facial-hits were rated higher (M. = 7.44, S.D. = .45) than medium-attractiveness facial-hits (M. = 4.83, S.D. = .25; $t\ (22) = 26.46$, $p < .01$; $d = 7.17$) and low-attractiveness facial-hits (M. = 3.58, S.D. = .53; $t\ (22) = 23.12$, $p < .01$; $d = 7.85$). Medium-attractiveness facial-hits were also rated higher for attractiveness ratings than low-attractiveness facial-hits ($p < .01$; $d = 3.02$). For emotion discrimination, high-attractiveness facial-misses (M. = 6.23, S.D. = .68) reported significantly higher attractiveness ratings compared to medium-attractiveness facial-misses (M. = 4.23, S.D. = 1.14; $t\ (15) = 4.91$, $p < .01$; $d = 2.13$) and low-attractiveness facial-misses (M. = 4.59, S.D. = .49; $t\ (11) = 9.44$, $p < .01$; $d = 2.77$).

These results suggested that the appraisal of attractiveness could not be performed in the absence of conscious awareness (face-detection misses) and that correct face detection (hits) was a necessary condition for the appraisal of attractiveness (see Figure 4). Interestingly, although correct emotion discrimination (hits) enhanced the acuity of the appraisal for attractiveness, incorrect emotion discrimination (misses) reported significant differences between different appearance types, suggesting that emotion discrimination was not necessary for the appraisal of attractiveness from faces.

Figure 4: Attractiveness Ratings for Hits and Misses

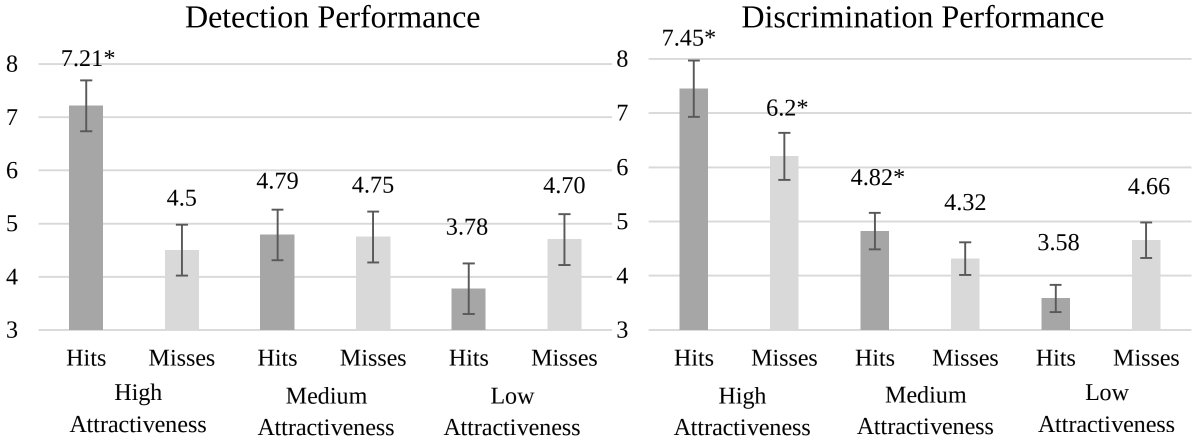


Figure 4: Mean Attractiveness ratings per Appearance Type for Hits and Misses for face-detection and emotion-discrimination performance in Study Two. Error bars indicate standard error of the mean (±2). Asterisks (*) indicate significance at p < .01 level.

Exploratory Analysis. The exploratory analysis in the current stage tested whether appearance type and emotion interact under conditions of backwards masking to influence attractiveness ratings. The analysis revealed a trend for significance for an interaction between Appearance Type and Emotion (F (4, 28) = 2.01, $p = .06$; $\eta^2 = .22$); no pairwise comparisons survived the Bonferonni corrections (Figure 5). These findings suggested that the interaction between emotion and attractiveness was associated with minor differences in attractiveness ratings. Further hits and misses analysis for this interaction could not be performed because of insufficient face-detection and emotion-discrimination hits and misses responses for several appearance type and emotion combinations (e.g. high-attractiveness angry faces, low-attractiveness neutral faces) and because the available responses did not meet ($P_{(1-\beta)} = .34$) the minimum statistical power requirement criteria ($P_{(1-\beta)} \geq .8$).

Figure 5: Attractiveness and Emotion under Conditions of Backwards Masking

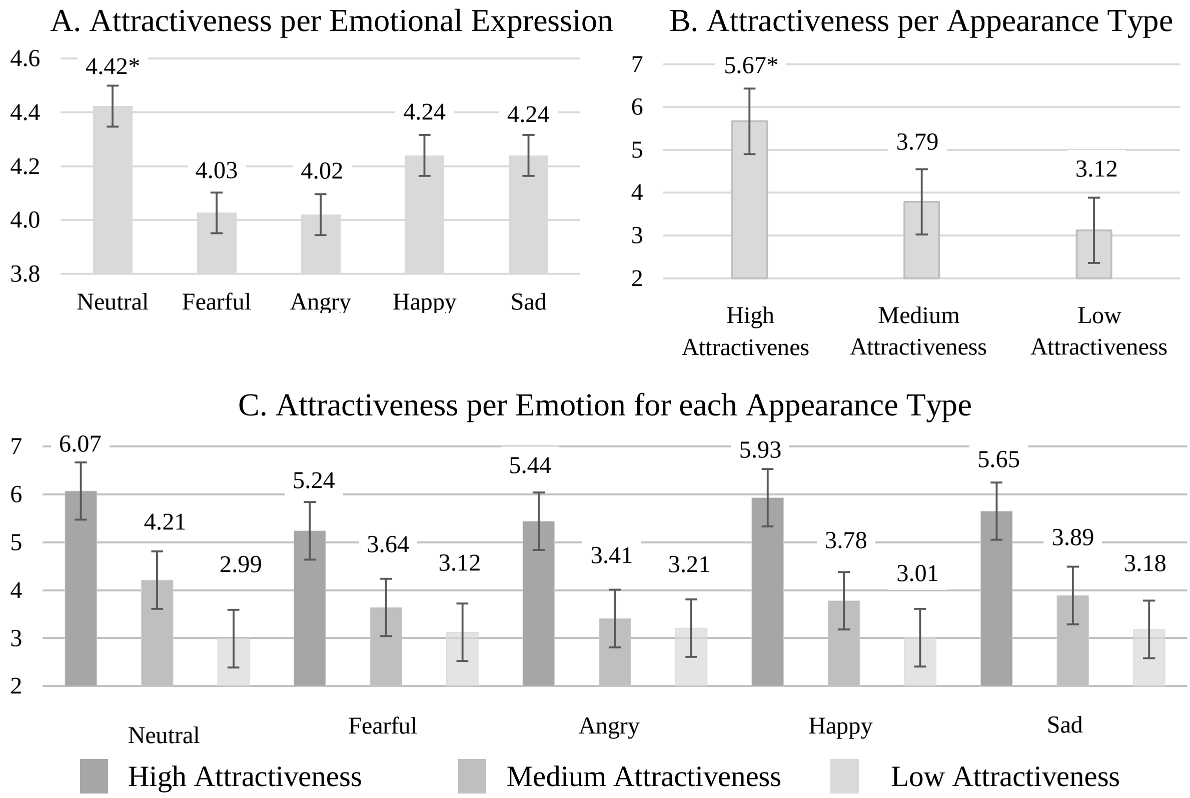


Figure 5: Attractiveness ratings for each Appearance type (high, medium and low attractiveness) for each Emotion (neutral, fearful, angry, happy and sad). A significance trend ($p = .06$) for an Appearance to Emotion Interaction was reported although no pairwise comparisons survived the Bonferonni corrections. Asterisks (*) indicate significance at $p < .01$ level.

Summary of Findings

The primary aims of the current manuscript was to explore whether attractiveness influences face-detection and emotion-discrimination performance and whether self-reports for the appraisal of facial attractiveness require conscious awareness. We implemented several methodological developments to explore these hypotheses such as extensive pilot experimental stimuli controls, signal detection analysis using sensitivity index A, Bayesian assessment for chance-level significance that would indicate stimuli invisibility, and analysis of hits and misses in face-detection and emotion-discrimination performance for attractiveness-rating responses. We found that attractiveness influences face-detection and emotion-discrimination performance, and more specifically that high-attractiveness faces can be detected and discriminated more accurately than other appearance types. Our analysis also revealed that specific appearance type and emotion combinations such as high-attractiveness fearful, angry and sad faces, and low-attractiveness neutral faces were detected and discriminated more accurately than other appearance type to emotion combinations. Critically, we found that face detection (hits) was a necessary condition for the appraisal of attractiveness for high-attractiveness faces and that when participants had absence of conscious awareness of the presented face (face-detection misses) they did not rate high-attractiveness faces higher than other appearance types. Interestingly, although correct emotion discrimination (hits) enhanced the acuity of the appraisal for high-attractiveness faces, incorrect emotion discrimination (misses) also reported higher ratings for high-attractiveness faces compared to other appearance types, suggesting that emotion discrimination is not necessary for the appraisal of attractiveness from faces. Finally, our exploratory analysis revealed that the interaction of emotion and appearance type impacts face-detection and emotion-discrimination performance but has only a minor effect in ratings for attractiveness.

General Discussion

Previous research has reported that attractiveness can be appraised from minimal information, such as pixelated images (Bachmann, 2007) and brief presentations (Olson & Marshuetz, 2005), and processed despite interocular suppression (Hung et al., 2016). For example, Olson and Marshuetz (2005) reported that high-attractiveness faces presented for as little as 13 ms, preceded by a high-frequency contour scrambled face mask for 39 ms and followed by a carton mask for 39 ms, were associated with higher attractiveness ratings than low-attractiveness faces presented under the same conditions. They also reported that 13 ms high-attractiveness faces elicited shorter reaction times for the appraisal of subsequently presented positively valanced words than low-attractiveness faces (Olson & Marshuetz, 2005;

p. 500). Similarly, Hung, Nieh and Hsieh (2016) ran a series of experiments using interocular suppression (Sengpiel & Blakemore, 1995). They used staircase reduction of contrast visibility when bilaterally presenting high-attractiveness and low-attractiveness faces followed by Gabor patches. They reported a reduction in Gabor lines orientation discrimination performance for lateral high-attractiveness faces due to perceptual inhibition of return, suggesting that high-attractiveness faces were attended despite the staircase reduction in visual contrast (Klein, 2000). These findings have been used to propose that attractiveness can be appraised from minimal information and have also been used to suggest that attractiveness can be processed and reported in the absence of conscious awareness (see also Kleckner et al., 2018) .

The current findings refer to whether attractiveness can be appraised (Olson & Marshuetz, 2005) without conscious awareness. Our findings support previous research (Olson & Marshuetz, 2005; Bachmann, 2007; Hung et al., 2016) in that – like other evolutionary important stimuli (Mineka & Öhman, 2002; Öhman, 2009) that confer "sociobiological value" (Bachmann, 2007; p. 848) – high-attractiveness faces convey highly-salient cues (Fink & Penton-Voak, 2002) that render them more accurately detectable during brief presentations (Calvo & Lundqvist, 2008; Lähteenmäki, Hyönä, Koivisto & Nummenmaa, 2015). The current findings also add to the existing literature by suggesting that high-attractiveness faces are discriminated more accurately (Adolphs, 2008) and that correct discrimination of the expressed emotion enhances the acuity but is not a necessary condition for the appraisal of the attractiveness from faces (Sergent & Dehaene, 2004).

Despite this partial consensus with previous research, the current manuscript employed different methodological assessment and statistical applications (Tsikandilakis et al., 2017; Tsikandilakis & Chapman, 2018) compared to previous publications. These included the application of signal detection theory (Zhang & Mueller, 2004), Bayesian analysis of chance-level significance (Dienes, 2015) that would indicate stimuli invisibility (Erdelyi, 2004), and separate analysis for hits and misses for attractiveness ratings (Pessoa, 2005). Based on the current methods, our results also partly disagree with previous findings and we propose that the appraisal of attractiveness does require conscious target-detection meta-awareness of facial characteristics (Tsikandilakis et al., 2017). This finding does not *per se* oppose previous findings that have suggested that attractiveness can be processed in the absence of awareness (Hung et al., 2016) because ratings for attractiveness were explicitly measured in the current study. Instead, the key finding of the current report is that the appraisal of attractiveness cannot be reported in the absence of awareness and that correct detection (hits) of a masked face was a required condition for the appraisal of attractiveness. Therefore, we suggest that to make conscious judgements (see also Lau, 2008) in relation to the attractiveness of a presented face, conscious detection of that face is a necessary condition.

Finally, in respect to our exploratory analysis we reported that high-attractiveness angry, fearful and sad faces as well as low-attractiveness neutral faces were detected and discriminated more accurately than other appearance type to emotion combinations. We also reported a trend for an interaction between emotion and appearance type in relation to attractiveness ratings under conditions of backwards masking. These findings are a novel contribution to the field and support that high-attractiveness angry and fearful faces confer a face-detection superiority effect (Lundqvist, Bruce & Öhman, 2015) compared to other stimuli, possibly due to their high-salience and sociobiological value (Brooks et al., 2012). The finding that emotion and appearance type revealed only a trend for an interaction in relation to attractiveness ratings means that we could not provide solid support for the idea that particularly positive emotional expressions influence attractiveness ratings when these are presented under conditions of backwards masking (O'Doherty et al., 2002). This could relate to low power in the current experiment and/or relate to participants' inability to process and integrate multiple informational input when the signal strength is reduced under conditions of backwards masking (Baars, 2002).

Limitations

To our knowledge the current study is the first attempt to explore the appraisal of attractiveness using backwards masking and angry, fearful, sad and happy faces in addition to neutral faces that have been employed in previous research (Olson & Marshuetz, 2005; Hung et al., 2016). Despite previous research suggesting a happy face detection superiority effect during low level visual processing (Miyazawa & Iwasaki, 2010) and higher attractiveness ratings for happy faces during supraliminal presentations (O'Doherty et al., 2003), we were not able to find higher face-detection or emotion-discrimination responses for happy faces for any appearance type. Additionally, emotional expression and appearance type provided only an overall trend for an interaction with attractiveness ratings as the dependent variable. Future research could benefit from a dedicated and appropriately powered exploration of the interaction between emotional expressions and appearance types (Penton-Voak & Chang, 2008) under conditions of backwards masking, and particularly the exploration of the possibility that different emotional type (fearful, angry, happy, sad and neutral) to appearance type (high, medium and low attractiveness) combinations could result in differentiating patterns of attractiveness reports (Rhodes, 2006). In the same context, it is possible, although outside the scope of the current manuscript, that low-attractiveness faces could also report face-detection and emotion-discrimination differences compared to other appearance types, such as mid-attractiveness faces, because they confer health and fitness related perceptual cues with avoidance response value (Pazda, Thorstenson, Elliot, & Perrett, 2016; Jaeger, Wagemans, Evans, & van Beest, 2018). This possibility is also supported by the data in the current manuscript (see Table 1), and future research could benefit from a dedicated exploration of face-detection and emotion-discrimination performance of low-attractiveness faces under conditions of backwards masking.

Conclusions

The current study assessed whether attractiveness has an impact on face detection and emotion discrimination using backwards masking, and whether attractiveness ratings require conscious awareness. Our results revealed that high-attractiveness faces were detected and discriminated more accurately than other appearance types but also that detection of a face was a necessary condition for the appraisal of attractiveness. This is a novel contribution to the field and suggests that appraisal of attractiveness requires conscious awareness of a presented face.

Acknowledgements

The primary author would like to thank Sevim Hodge, Harry Purser, Martyn Quigley, Alison Mostyn, Jon Peirce, Jan Derfuss and Anthony Watkins for their contribution in this project personally and academically. The current research has no conflicts of interest. The current research project was funded by the Economic and Social Research Council of the university of Nottingham. The data for the current report have been made open access at https://osf.io/qh7z4/.

References

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current opinion in neurobiology*, 18(2), 166-172.

Alexithymia. (2017, September 17). *Alexithymia: Emotional blindness*. Retrieved from http://www.alexithymia.us/test-alex.html.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. Trends in cognitive sciences, 6(1), 47-52.

Bachmann, T. (2007). When beauty breaks down: Investigation of the effect of spatial quantisation on aesthetic evaluation of facial images. *Perception, 36*(6), 840-849.

Bachmann, T., & Francis, G. (2013). *Visual masking: Studying perception, attention, and consciousness*. Academic Press.

Barber, N. (1995). The evolutionary psychology of physical attractiveness: Sexual selection and human morphology. *Ethology and Sociobiology*, *16*(5), 395-424.

Berry, D. A. (1996). Statistics: A Bayesian perspective, London, England: Duxbury Press.

Brooks, S. J., Savov, V., Allzén, E., Benedict, C., Fredriksson, R., & Schiöth, H. B. (2012). Exposure to subliminal arousing stimuli induces robust activation in the amygdala, hippocampus, anterior cingulate, insular cortex and primary visual cortex: a systematic metaanalysis of fMRI studies. *NeuroImage*, *59*(3), 2962-2973.

Cai, Z., Hahn, A. C., Zhang, W., Holzleitner, I. J., Lee, A. J., DeBruine, L. M., & Jones, B. C. (2018, September 12). *No evidence that women's facial attractiveness, femininity, or averageness are valid health cues*. Retrieved from https://osf.io/f9tu2/

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior research methods*, *40*(1), 109-115.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends in cognitive sciences, 10(5), 204-211

Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. *Behavioural methods in consciousness research*, *2015*, 199-220.

Erdelyi, M. H. (2004). Subliminal perception and its cognates: Theory, indeterminacy, and time. *Consciousness and Cognition*, *13*(1), 73-91.

Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science*, *11*(5), 154-158.

Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2006). Facial symmetry and judgements of attractiveness, health and personality. Personality and Individual Differences, 41(3), 491499.

Fisher, R. J., & Ma, Y. (2014). The price of being beautiful: Negative effects of attractiveness on empathy for children in need. *Journal of Consumer Research*, *41*(2), 436-450.

Foo, Y. Z., Simmons, L. W., & Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Scientific Reports*, 7, 1–12

Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and income for men and women in management. *Journal of Applied Social Psychology*, *21*(13), 1039-1057.

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of comparative psychology*, *108*(3), 233.

Griffin, A. M., & Langlois, J. H. (2006). Stereotype Directionality and Attractiveness Stereotyping: Is Beauty Good or is Ugly Bad? *Social Cognition*, *24*(2), 187–206.

Grimshaw, G. M., Bulman-Fleming, M. B., & Ngo, C. (2004). A signal-detection analysis of sex differences in the perception of emotional faces. *Brain and Cognition*, *54*(3), 248-250.

Gur, R.C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., Turner, T., Bajcsy, R., Posner, A. and Gur, R.E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of neuroscience methods*, *115*(2), 137-143.

Hickie, I. B., Davenport, T. A., Hadzi-Pavlovic, D., Koschera, A., Naismith, S. L., Scott, E. M., & Wilhelm, K. A. (2001). Development of a simple screening tool for common mental disorders in general practice. *The Medical journal of Australia*, *175,* S10-7.

Hong, S., Tandoc Jr, E., Kim, E. A., Kim, B., & Wise, K. (2012). The real you? The role of visual cues and comment congruence in perceptions of social attractiveness from Facebook profiles. *Cyberpsychology, Behavior, and Social Networking*, 15(7), 339-344.

Hung, S. M., Nieh, C. H., & Hsieh, P. J. (2016). Unconscious processing of facial attractiveness: Invisible attractive faces orient visual attention. *Scientific reports*, *6*, 37117.

Jaeger, B., Wagemans, F. M., Evans, A. M., & van Beest, I. (2018). Effects of facial skin smoothness and blemishes on trait impressions. Perception, 0301006618767258.

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Lao, J., ... & DeBruine, L. M. (2018). No evidence that more physically attractive women have higher estradiol or progesterone. bioRxiv, 136515.

Klein, R. M. (2000). Inhibition of return. *Trends in cognitive sciences*, 4(4), 138-147.

Kleckner, I. R., Anderson, E. C., Betz, N. J., Wormwood, J. B., Eskew, R. T., & Barrett, L. F. (2018). Conscious awareness is necessary for affective faces to influence social judgments. Journal of Experimental Social Psychology, 79, 181–187.

Lähteenmäki, M., Hyönä, J., Koivisto, M., & Nummenmaa, L. (2015). Affective processing requires awareness. *Journal of Experimental Psychology: General*, *144*(2), 339

Lander, K., Bruce, V. and Hill, H. (2001), Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology,* 15: 101116.

Langlois, J. H., & Stephan, C. W. (1981). Beauty and the beast: The role of physical attractiveness in the development of peer relations and social behavior. *Developmental social psychology: Theory and research*, 152-168.

Lau, H. (2008). “Are we studying consciousness yet?” in *Frontiers of Consciousness: Chichele Lectures*, eds L. Weiskrantz and M. Davies, Oxford: Oxford University Press, 245– 258.

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1638–1659.

Lyons, M. J., Campbell, R., Plante, A., Coleman, M., Kamachi, M., & Akamatsu, S. (2000). The Noh mask effect: vertical viewpoint dependence of facial expression perception. *Proceedings of the Royal Society of London B: Biological Sciences*, *267*(1459), 2239-2245.

Lundqvist, D., Bruce, N., & Öhman, A. (2015). Finding an emotional face in a crowd: Emotional and perceptual stimulus factors influence visual search efficiency. Cognition and Emotion, 29(4), 621-633.

Massar, K., & Buunk, A. P. (2010). Judging a book by its cover: Jealousy after subliminal priming with attractive and unattractive faces. *Personality and individual differences*, *49*(6), 634-638.

Mineka, S. and Öhman, A. (2002). Phobias and preparedness: the selective, automatic, and encapsulated nature of fear. *Biological Psychiatry*, 52(10), pp. 927–937.

Miyazawa, S., & Iwasaki, S. (2010). Do happy faces capture attention? The happiness superiority effect in attentional blink. *Emotion, 10*(5), 712-716.

Noldus. (2018, June 3). Reference manual. http://www.noldus.com/human-behaviorresearch/products/facereader

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, *41*(2), 147-155.

Öhman, A. (2009). Of snakes and faces: An evolutionary perspective on the psychology of fear. *Scandinavian journal of psychology*, *50*(6), 543-552.

Olson, I. R., & Marshuetz, C. (2005). Facial Attractiveness Is Appraised in a Glance. *Emotion, 5*(4), 498-502.

Pazda, A. D., Thorstenson, C. A., Elliot, A. J., & Perrett, D. I. (2016). Women's facial redness increases their perceived attractiveness: Mediation through perceived healthiness. Perception, 45(7), 739-754.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of neuroscience methods, 162(1-2), 8-13.

Penton-Voak, I. S., & Chang, H. Y. (2008). Attractiveness judgements of individuals vary across emotional expression and movement conditions. *Journal of Evolutionary Psychology, 6*(2), 89-100.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness?. *Current opinion in neurobiology*, *15*(2), 188-196.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.*, *57*, 199-226.

Ritchie, K. L., Palermo, R., & Rhodes, G. (2017). Forming impressions of facial attractiveness is mandatory. *Scientific Reports*, 7(469), 1–8.

Roberts, S.C., Little, A.C., Gosling, L.M., Perrett, D.I., Carter, V., Jones, B.C., Penton-Voak, I. & Petrie, M. (2005). MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior*, *26*(3), 213-226.

Saegusa, C., & Watanabe, K. (2016). Judgments of facial attractiveness as a combination of facial parts information over time: Social and aesthetic factors. Journal of Experimental Psychology: Human Perception and Performance, 42(2), 173.

Saxton, T. K., Caryl, P. G., & Roberts, S. C. (2006). Vocal and facial attractiveness judgments of children, adolescents and adults: the ontogeny of mate choice. *Ethology*, *112*(12), 1179-1185.

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*(1431), 1913-1917.

Scott, I. M. L., Clark, A. P., Boothroyd, L. G., & Penton-Voak, I. S. (2013). Do men’s faces really signal heritable immunocompetence? *Behavioral Ecology*, 24(3), 579–589.

Sengpiel, F., Blakemore, C., & Harrad, R. (1995). Interocular suppression in the primary visual cortex: a possible neural basis of binocular rivalry. Vision research, 35(2), 179-195.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological science*, *15*(11), 720-728.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust?. Methodology.

Smith, M.L., Perrett, D.I., Jones, B.C., Cornwell, R.E., Moore, F.R., Feinberg, D.R., Boothroyd, L.G., Durrani, S.J., Stirrat, M.R., Whiten, S. and Pitman, R.M. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1583), 135-140.

Soper, D., (2018, September 17). Type II-Error. Retrieved from https://www.danielsoper.com/statcalc/calculator.aspx?id=50.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.

Swami, V., & Furnham, A. (2008). The psychology of physical attraction. Hove: Psychology Press.

Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness. Sexual nature, sexual culture, 80-118.

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental science*, *10*(5), 547-558.

Tsikandilakis, M., Chapman, P., & Peirce, J. (2017). Target meta-awareness is a necessary condition for physiological responses to masked emotional faces: Evidence from combined skin conductance and heart rate assessment. *Consciousness and cognition*.

Tsikandilakis, M., & Chapman, P. (2018). Skin conductance responses to masked emotional faces are modulated by hit rate but not signal detection theory adjustments for subjective differences in the detection threshold. *Perception*.

van der Ploeg, M. M., Brosschot, J. F., Versluis, A., & Verkuil, B. (2017). Peripheral physiological responses to subliminally presented negative affective stimuli: A systematic review. *Biological Psychology, 129*, 131-153.

Werheid, K., Schacht, A., & Sommer, W. (2007). Facial attractiveness modulates early and late event-related brain potentials. *Biological psychology*, *76*(1-2), 100-108.

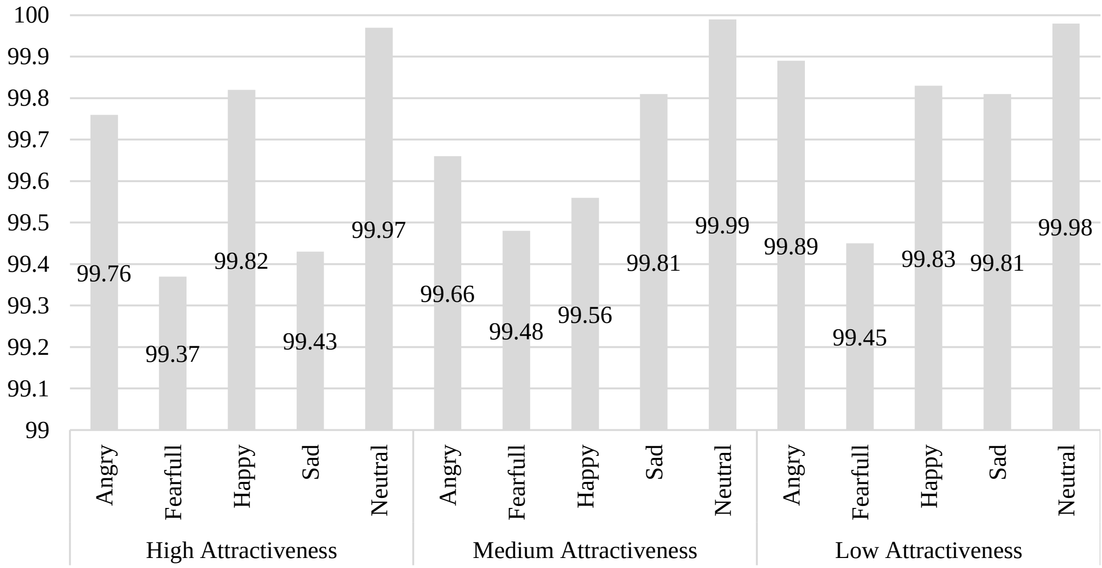
Wessa P., (2016), Testing Mean with known Variance (type II error) (v1.0.5) in Free Statistics Software (v1.2.1), Office for Research Development and Education, URL http://www.wessa.net/rwasp_hypothesismean3.wasp/

Winkler, C., & Rhodes, G. (2005). Perceptual adaptation affects attractiveness of female bodies. British Journal of Psychology, 96(2), 141-154.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203-212.

Appendix

1.1: Noldus Emotional Recognition Certainty Scores (%)



2.1: Gender Effects

In accordance with previous research that used frequentist statistics (Olson & Marshuetz, 2005) we were not able to report actor or participant related gender effects in our analysis. This is a counter-intuitive and yet recurrently reported finding in the current research area (Little et al., 2011). Future research could benefit from an assessment of the possible effects of facial cropping techniques (Lyons et al., 2000) and the possible role of bodily characteristics (Barber, 2005; Winkler & Rhodes, 2005) in exploring gender-specific differences in attractiveness ratings.

3.1: Bayesian Analysis for Attractiveness in Stage one for Gender

A Bayesian analysis (Dienes, 2015) with corrected degrees of freedom (df < 30; $SE = (SE\ x \left(1 + \left(\frac{20}{df x df}\right)\right)$) (Berry, 1996) revealed that for attractiveness ratings in stage one male (S.E. = .07; B = .09) and female (S.E. = .07; B = .09) participant responses were significantly within the predefined uniform credible intervals (4 < M. > 6).

4.1 Bayesian Analysis of Emotionality in Stage Two between Appearance Types

A Bayesian analysis (Dienes, 2015) with corrected degrees of freedom (df < 30; $SE = (SE\ x \left(1 + \left(\frac{20}{df x df}\right)\right)$) (Berry, 1996) revealed that in stage two emotionality ratings for high-attractiveness faces (S.E. = .05; B = .08), medium-attractiveness faces (S.E. = .04; B = .06) and low-attractiveness faces (S.E. = .04; B = .07) were significantly within the predefined uniform credible intervals (5 > M. < 7).

5.1: Bayesian analysis for Emotionality in Stage Two for Gender

A Bayesian analysis (Dienes, 2015) with corrected degrees of freedom (df < 30; $SE = (SE\ x\ \left(1\ +\ \left(\frac{20}{df x df}\right)\right))$ (Berry, 1996) revealed that in stage two emotionality ratings for male (S.E. = .03; B = .07) and female (S.E. = .03; B = .07) were significantly within the predefined credible intervals (5 > M. < 7).

6.1: Bayesian Analysis in Stage Three for Visual Contrast

A Bayesian analysis (Dienes, 2015) with corrected degrees of freedom (df < 30; $SE = (SE\ x\ \left(1\ +\ \left(\frac{20}{df x df}\right)\right))$ (Berry, 1996) revealed that non-facial blurs (S.E. = .13; B = .16) and facial stimuli (S.E. = .06; B = .08) were significantly within the predefined uniform credible intervals (4 > M. < 6).