Perception, 2015, volume 44, pages 79–86

doi:10.1068/p7904

SHORT REPORT

Facial cosmetics have little effect on attractiveness judgments compared with identity

Alex L Jones[1,2], Robin S S Kramer[3]§
[1] Department of Psychology, Gettysburg College, Gettysburg, PA 17325, USA; [2] School of Psychology, Bangor University, Bangor LL57 2AS, Gwynedd, UK; [3] School of Psychology, University of Aberdeen, Aberdeen AB24 3FX, UK; e-mail: remarknibor@gmail.com
Received 30 October 2014, in revised form 8 December 2014

Abstract. The vast majority of women in modern societies use facial cosmetics, which modify facial cues to attractiveness. However, the size of this increase remains unclear—how much more attractive are individuals after an application of cosmetics? Here, we utilised a 'new statistics' approach, calculating the effect size of cosmetics on attractiveness using a within-subjects design, and compared this with the effect size due to identity—that is, the inherent differences in attractiveness between people. Women were photographed with and without cosmetics, and these images were rated for attractiveness by a second group of participants. The proportion of variance in attractiveness explained by identity was much greater than the variance within models due to cosmetics. This result was unchanged after statistically controlling for the perceived amount of cosmetics that each model used. Although cosmetics increase attractiveness, the effect is small, and the benefits of cosmetics may be inflated in everyday thinking.

Keywords: facial cosmetics, attractiveness, effect size, social perception

1 Introduction

Decoration of the body is one of the most common human behaviours, present across different cultures and throughout history (Jablonski, 2006). In modern societies the use of cosmetics by females to alter facial appearance is nearly universal. Over 80% of women over the age of 18 years wear cosmetics (Etcoff, 1999), and the value of the global cosmetics industry exceeds €130 billion (Rossi, Prlic, & Hoffman, 2007).

Cosmetics can influence perceptions of social traits, with facial attractiveness perhaps being the most studied of these. Perceptions of attractiveness increase with an application of cosmetics (Cash, Dawson, Davis, Bowen, & Galumbeck, 1989; Etcoff, Stock, Haley, Vickery, & House, 2011; Mulhern, Fieldman, Hussey, Lévêque, & Pineau, 2003), and when wearing cosmetics, females provide higher estimates of their own attractiveness (Cash et al., 1989). An application of cosmetics also increases perceptions of traits related to attractiveness, with wearers perceived as healthier and from a higher socioeconomic background (Nash, Fieldman, Hussey, Lévêque, & Pineau, 2006). Faces with cosmetics even induce greater activation in the reward centres of the brain, such as the medial orbitofrontal cortex (Ueno et al., 2014).

How do cosmetics increase facial attractiveness? The contrast between features and skin is naturally higher in female faces, and this facial contrast correlates positively with perceived femininity (Russell, 2009). A typical application of cosmetics darkens facial features and lightens the skin, increasing this contrast (Russell, 2010), as well as masking age-related declines in these contrasts (Porcheron, Mauger, & Russell, 2013). Cosmetics also accentuate attractive colouration of the lips (Stephen & McKeegan, 2010), and homogenise skin texture, removing blemishes or uneven colour distributions (Samson, Fink, & Matts, 2010).

§ Corresponding author.

Said and Todorov (2011) demonstrated that attractive female faces have lighter skin, darker eyes, and redder lips than the average female—all components of skin colouration that cosmetics alter. Indeed, Kościński (2012) demonstrated that the quantities of cosmetics applied to the eyes and mouth were predictors of facial attractiveness, with more makeup producing higher attractiveness ratings. Of these, the amount of eye makeup was a particularly strong predictor (see also Mulhern et al., 2003).

Female facial appearance without cosmetics follows established rules regarding perceived attractiveness, with more symmetrical, average, and feminine faces rated as more attractive (Rhodes, 2006). These properties are relatively unchangeable, being fixed characteristics of facial structure. However, emotional expressions can alter attractiveness judgments (Tracy & Beall, 2011). Importantly, though, Morrison, Morris, and Bard (2013) found that facial identity (fixed, unchangeable aspects of the face—between-person variation) explained more than twice the amount of variance in attractiveness judgments as facial expressions did. These results suggest that attractiveness is a stable property of the face, with even positive social signals like expressions doing relatively little to alter perceptions. However, emotional expressions are fleeting, occurring over short time scales, while cosmetics produce a lasting and direct manipulation of attractive facial properties. We therefore ask whether cosmetics also have only minor effects on attractiveness perceptions, since they do not alter facial structure, or whether they produce more substantial changes, given their longer-lasting and targeted nature.

In the present study, we examine how much an application of cosmetics contributes to perceptions of attractiveness beyond differences simply due to an individual's identity. We employ a 'new statistics' approach (Cumming, 2014), whereby within-face variation in attractiveness, due to cosmetics, is compared with between-face variation, due to differences in facial identity. This involves a shift in focus from unreliable *p*-values that answer essentially dichotomous questions ("do cosmetics have a significant effect on attractiveness?") to an examination of effect sizes ("how large is the effect of cosmetics on attractiveness?"). This approach facilitates a richer way of examining data, reframing questions in terms of quantifying effects rather than simply declaring their presence or absence (eg Swami, Tovée, & Harris, 2013).

In order to compare the effect sizes of cosmetics and identity, we collected attractiveness ratings for photographs of the same models with and without cosmetics. Each rater saw only one image of each model (either with or without cosmetics, selected randomly), avoiding carryover effects while allowing both factors to vary simultaneously. In this way, we could address straightforward but novel questions about cosmetics use—how much more attractive do cosmetics make someone, and do they overcome inherent differences between individuals?

2 Results

2.1 *Analysis of variance*

Each image received an average of 31.00 ratings (SD = 6.74). To examine agreement in ratings, we calculated the variance for ratings of each image, and averaged these to provide a measure of spread for each cosmetics condition separately: without cosmetics, SD = 1.07; with cosmetics, SD = 1.17. Note that ratings were given using a 7-point scale. As such, the relatively low variability indicates general agreement in ratings for each image (in line with previous research, eg Coetzee, Greeff, Stephen, & Perrett, 2014). In order to examine the potential effect of rater sex (differences in male and female perceptions of attractiveness), we split the data into male and female raters before averaging the ratings for each image. This gave each model an average attractiveness rating in each cosmetics condition, for each sex of rater. All subsequent analyses used the model as the unit of analysis.

To obtain the η^2 effect sizes, we first calculated the sums of squares (SS) for each factor and interaction using an analysis of variance. Our data followed a repeated-measures design, where each model had an attractiveness rating (by averaging across raters, described above) for both levels of cosmetics condition (with and without) and rater sex (male and female). Although often of little importance in research, we specifically included consideration of the variance explained by 'identity'—that is, differences between individuals (in this case, models). Therefore, this factor and its interactions with other variables were included in the full analysis[1] summarised in table 1.

Table 1. Results of the analysis of variance.

Source	df	SS	η^2	F	p
Identity (I)	43	88.59	0.69		
Rater sex	1	18.19	0.14	165.15	<0.001
Rater sex × I	43	4.74	0.04		
Cosmetics	1	2.77	0.02	11.46	0.002
Cosmetics × I	43	10.39	0.08		
Rater sex × Cosmetics	1	0.02	<0.001	0.21	0.65
Rater sex × Cosmetics × I	43	4.52	0.04		
Total	175	129.23			

Note. There is no appropriate error term for testing effects involving differences between models. As such, there is no F ratio calculated for these terms. df = degrees of freedom, SS = sums of squares. η^2 values were obtained by dividing the source SS by the total SS.

The effect of rater sex was the result of males [$M = 2.68$, 95% CI (2.49, 2.87)] providing lower ratings than females [$M = 3.32$, (3.07, 3.57)]. This pattern has been previously demonstrated in the literature (eg Cross & Cross, 1971) and is of little interest here. In addition, faces with cosmetics [$M = 3.13$ (2.88, 3.37)] were given higher ratings of attractiveness than those without [$M = 2.88$ (2.66, 3.09)]. However, as table 1 illustrates, the effect size of identity was 34.5 times larger than this cosmetics effect, which indicates that differences in attractiveness between individuals explain a great deal more variance than cosmetics, which explain relatively little.

As table 1 shows, the effect size of the interaction between identity and cosmetics ($\eta^2 = 0.08$), although relatively small, suggests that the application of cosmetics affected the attractiveness of each model differently. Indeed, this may be expected since our cosmetics manipulation provided no restrictions on how much makeup each model applied. Could differences in the amount of makeup between models explain why the effect of identity is much larger than the effect of cosmetics?

2.2 *Analysis of covariance*

Even though models applied the same range of cosmetics themselves, each individual likely had their own history of using makeup, individual skill level, and preferences for a particular appearance, leading to differences in the amount of makeup applied. Might these differences inflate the effect of identity?

[1] This full analysis allows the 'identity' term to interact with the other variables, accounting for the fact that different treatments may affect different models in different ways. As such, it is usually preferred over the reduced version where these interactions are not included (Howell, 1997). As a result of the inclusion of 'identity' and its interactions, we have no residuals left for estimating error. However, the SS for the interaction between identity and each variable can be used as an error estimate for that variable (Howell, 1997, page 487).

To address this possibility, we presented ten new participants (age $M = 22.50$ years, SD = 6.82 years, three males) with both the 'without' and 'with cosmetics' photographs of each model on screen next to each other, and asked "how much makeup has this person put on?" Participants indicated their response on a 1 (very light) to 7 (very heavy) scale, and we averaged these ratings for each model to provide a 'cosmetics quantity' score. We then repeated our analysis as above, but entered this 'quantity' score as a covariate after mean centring the variable ($M = 3.68$, SD = 0.88; see Delaney & Maxwell, 1981). In analyses of covariance, the addition of the covariate adjusts the SS attributable to each source of variation from the original analysis, which in turn adjusts the η^2 effect sizes. If the differences in the way cosmetics were applied by different models artificially inflated differences between individuals, then we should see a reduction in the size of the identity effect and a possible increase in the effect size for cosmetics.

With the introduction of the covariate, the effect size due to the amount of cosmetics applied ('quantity') was small ($\eta^2 = 0.01$), as were the effect sizes of its interactions (all $\eta^2 < 0.006$). The effect size for identity was slightly reduced ($\eta^2 = 0.67$), while the effect size due to cosmetics condition remained the same ($\eta^2 = 0.02$). Therefore, even adjusting for differences in individual cosmetics use, identity was still 31.4 times more important than cosmetics.

Finally, we examined the effect of the quantity of cosmetics on attractiveness judgments directly. We calculated the difference in ratings between cosmetics conditions, and correlated these difference scores with the perceived amount of makeup applied by each model. The quantity of cosmetics showed no relationship with the attractiveness changes for models as rated by women [$r_{42} = –0.12$, 95% CI (–0.40, 0.18), $p = 0.43$] or by men [$r_{42} = 0.03$, (–0.26, 0.33), $p = 0.81$]. Therefore, the amount of cosmetics applied by models did not predict the resulting change in attractiveness, confirming our earlier analyses.

3 Discussion

In line with previous research (Cash et al., 1989), we find that cosmetics increased the attractiveness of our models. However, for the first time, we demonstrate that the difference in attractiveness ratings within models, due to cosmetics, explained only 2% of the total variation in ratings. In comparison, the differences between the models, due to differences in individual identity, explained 69%. This sizable difference was marginally reduced, but still present, when we statistically controlled for individual differences in the amount of cosmetics used.

Previous research has demonstrated the importance of surface cues in judgments of attractiveness (O'Toole, Price, Vetter, Bartlett, & Blanz, 1999) and the important role of skin texture in these perceptions. Why does an application of cosmetics, which demonstrably alters these properties, contribute so little to perceptions of attractiveness? It might be that cosmetics are not particularly effective at masking imperfections in skin texture or colouration, or that other cues to attractiveness, such as symmetry, are not well corrected (Korichi, Pelle-de-Queral, Gazano, & Aubert, 2011). The evidence here suggests that any benefits cosmetics do convey are very small in comparison with the natural distribution in attractiveness of faces in the population.

Here, we find no relationship between the amount of cosmetics our models applied and the resulting change in their perceived attractiveness. However, previous research (Kościński, 2012) found that the quantities of eye and lip makeup worn by students to their classes were significant predictors of attractiveness. There may be several reasons for this apparent contradiction. First, the quantity of cosmetics in the present study likely featured less variability. Our models were instructed to apply cosmetics for a 'night out',

which constrained the style of makeup used. Indeed, the range for the judged amount of cosmetics applied (SD = 0.88 on a 7-point scale) suggests models wore similar amounts. In comparison, Kościński sampled students who could have worn any amount (or even no makeup). Second, using the same models as in this study, we have previously demonstrated that cosmetics may increase attractiveness up to a point, beyond which additional cosmetics can produce detrimental effects (Jones, Kramer, & Ward, 2014). This might explain the lack of a straightforward association between cosmetics quantity and attractiveness. Third, the amount of cosmetics worn by our models was judged by displaying 'with' and 'without' images side-by-side for our raters to compare. In contrast, Kościński's raters provided judgments without this baseline for comparison. As such, perceptions of apparent cosmetics use could have been conflated with each individual's appearance. For example, it would be more difficult to accurately judge the amount of mascara worn without a 'no mascara' image for reference. Finally, cultural differences may explain the discrepancy. The present study (UK) and Kościński's (Poland) were carried out in different countries, where preferences for cosmetic styles and intensities may differ. Indeed, the influence of cosmetics on attractiveness may differ worldwide, although historical records (Corson, 1972) and contemporary practices (Russell, 2010) suggest consistencies. While these methodological differences between the two studies may explain the contrast in findings here, we encourage further research in order to properly address how the amount of makeup affects the change in attractiveness of wearers.

In the current paper we label differences in attractiveness between individuals as 'identity', and use single, passport-style images of individuals as a representation of their appearance. However, individuals vary greatly across different photographs, and this within-person variability has a significant impact on social judgments including attractiveness (Jenkins, White, Van Montfort, & Burton, 2011). Here, our assessment of the effect size of identity is limited to the differences between photographs taken under controlled conditions where we equate all other factors (lighting, pose, etc), and so future research might also consider cosmetic effects in comparison with the effect size due to within-person variability.

In the real world, different styles of cosmetics are frequently employed, and here we examine only one. Furthermore, cosmetics that are professionally applied (eg Etcoff et al., 2011) might lead to larger cosmetic effect sizes. However, such results would tell us less regarding the role of cosmetics in everyday life, our goal in the current study. In addition, our models were in their early twenties, as were the raters we recruited. While many social interactions and mating decisions involve young people of a similar age, we encourage future research to consider cosmetic effects in middle-aged and older adults. Equally, the sizes of cosmetic effects in other ethnicities have yet to be investigated and may well differ from the findings presented here.

One issue to consider is that, although we asked our models to pose with a neutral expression, an application of cosmetics may have resulted in subtle positive expression changes (through increased confidence or self-esteem; see Miller & Cox, 1982; Mulhern et al., 2003). As a result, slight smiles could increase attractiveness (Morrison et al., 2013). However, if our cosmetics condition also included a benefit due to smiling, then this would only increase the effect size due to our manipulation. Therefore, the importance of cosmetics reported here may be an overestimate of the effect due to cosmetics alone. Similarly, the attractiveness ratings of our models were, on average, generally low on the 1–7 scale. If the variability in attractiveness ratings of the models was increased (through a larger or different sample), then identity would account for more variance. Therefore, the effect of identity measured here may be a conservative estimate.

That cosmetics make a face more attractive is unsurprising. However, by examining effect sizes, we have shown that this increase is relatively small. While many cosmetics seem designed specifically and directly to increase attractiveness (Russell, 2010), they remain far less important compared with underlying facial appearance. Previous research has shown that women tend to apply more cosmetics than is optimal for an attractive appearance (Jones et al., 2014). Here, we extend this by demonstrating that those cosmetics that are worn result in only small alterations to our attractiveness. We also highlight a new approach to quantifying the importance of particular factors by suggesting a comparison between the size of their effects and those due to simple differences between people.

4 Methods

4.1 *Stimuli*

Models were recruited through advertisements and word of mouth at a Welsh university two years prior to the commencement of the present study. All experiments were carried out under approval of the university's Institutional Review Board, and both models and raters (see below) provided informed consent prior to participation (in accordance with the World Medical Association Helsinki Declaration as revised in October 2008). Each model was paid £6 for their participation. Forty-four self-reported white females (age $M = 21.18$ years, SD = 1.94 years) posed as models. We asked them to remove any facial jewellery as well as to thoroughly clean their face of all cosmetic products (necessary supplies were provided). Models also tied back their hair so as not to obscure their faces. Individuals were asked to pose with a neutral expression, and photographs were taken using a Nikon D3000 SLR camera mounted on a tripod, at a distance of approximately one metre. A white background was used, along with a Nikon SS-400 flash angled 45° towards the ceiling. Each model was photographed three times, and we used the clearest exposure as our final stimulus.

After the initial photograph, models were provided with a range of best-selling foundations, lipsticks, mascaras, eye shadows, eyeliners, brow pencils, and blushers, and were instructed to apply their cosmetics as though they were going on a 'night out'. They were allowed as much time as they needed to do this, and were left alone during this time. Finally, models were photographed again to capture their appearance with cosmetics (again with a neutral expression). Between shots, camera settings were kept constant, including lens aperture (F5.3), exposure time ($1/60$ s), and ISO speed rating (200). Afterwards, participants removed their cosmetics completely before taking part in a second study, which is not reported here.

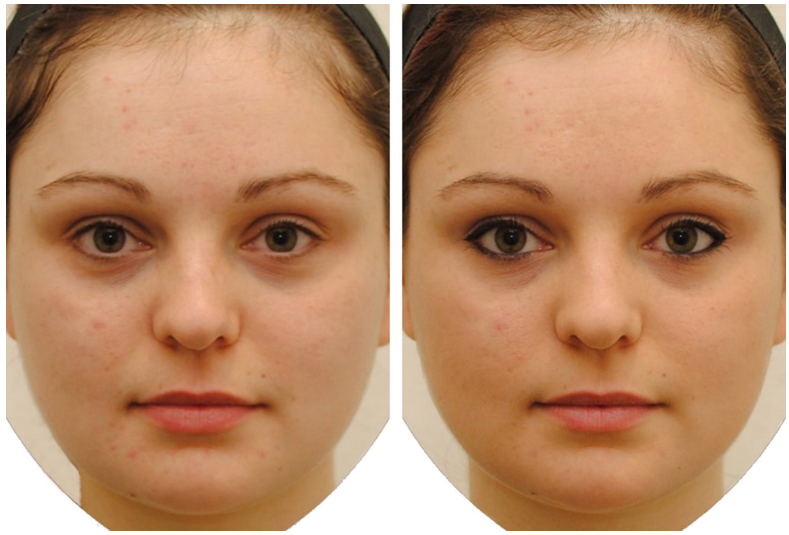
All photographs were subsequently rotated so the pupils lay along the same transverse plane, and were cropped to just above the hairline, to below the chin, and to just outside the widest points of the face (the zygions). The lower corners of the images were also cropped in order to remove or minimise the visibility of clothing. Finally, all images were scaled to the same height (550 pixels). Figure 1 provides an example of the final stimuli.

4.2 *Participants*

Sixty-two white participants volunteered to complete the study (age $M = 23.02$ years, SD = 3.82 years; thirty-one males), eighteen of which were university students in Wales (age $M = 23.16$ years, SD = 3.09 years; six males), while the rest were students and staff at a university in Scotland (age $M = 22.95$ years, SD = 4.11 years; twenty-five males).

4.3 *Procedure*

Participants rated the models for attractiveness on laptop computers using custom MATLAB software. Images were shown in a random order, with each model rated only once (44 trials in total) in a randomly chosen cosmetics condition (ie either with or without), preventing carryover effects (Morrison et al., 2013). If the same participant viewed models in both conditions, the influence of one image's rating on the other would be unavoidable. In contrast,



(a) (b)

Figure 1. [In colour online, see http://dx.doi.org/10.1068/p7904] An example of a model with (a) no cosmetics and (b) self-applied cosmetics for a night out. (These images are reproduced with permission of the model.)

if participants rated images in only one condition (all women wearing cosmetics, for example), this artificially exaggerates the differences between models due to identity since this is the only factor that would vary. Finally, our aim is to examine the relative sizes of the effects of cosmetics and facial identity, and so it is necessary to vary both for any given observer.

Images were rated on a scale of 1 (very unattractive) to 7 (very attractive). As the models were recruited from the same Welsh university as some of the participants, raters from this institution were also given a 'recognise' option onscreen, rather than providing a rating, and were instructed to use this for trials where they recognised the models. For these raters, an average of 0.73 trials (SD = 2.35 trials) were skipped. As such, ratings were collected only for photographs in which the model was not recognised by the rater.

References

Cash, T. F., Dawson, K., Davis, P., Bowen, M., & Galumbeck, C. (1989). Effects of cosmetics use on the physical attractiveness and body image of American college women. *Journal of Social Psychology*, **129**, 349–355.

Coetzee, V., Greeff, J. M., Stephen, I. D., & Perrett, D. I. (2014). Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender. *PLoS ONE*, **9**(7): e99629.

Corson, R. (1972). *Fashions in makeup: From ancient to modern times.* London: Peter Owen.

Cross, J. F., & Cross, J. (1971). Age, sex, race, and the perception of facial beauty. *Developmental Psychology*, **5**, 433–439.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, **25**, 7–29.

Delaney, H. D., & Maxwell, S. E. (1981). On using analysis of covariance in repeated measures designs. *Multivariate Behavioral Research*, **16**, 105–123.

Etcoff, N. (1999). *Survival of the prettiest.* New York: Doubleday.

Etcoff, N. L., Stock, S., Haley, L. E., Vickery, S. A., & House, D. M. (2011). Cosmetics as a feature of the extended human phenotype: Modulation of the perception of biologically important facial signals. *PLoS ONE*, **6**(10): e25656.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth.

Jablonski, N. G. (2006). *Skin: A natural history*. Berkeley, CA: University of California Press.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, **121**, 313–323.

Jones, A. L., Kramer, R. S. S., & Ward, R. (2014). Miscalibrations in judgements of attractiveness with cosmetics. *Quarterly Journal of Experimental Psychology*, **67**, 2060–2068.

Korichi, R., Pelle-de-Queral, D., Gazano, G., & Aubert, A. (2011). Relation between facial morphology, personality and the functions of facial make-up in women. *International Journal of Cosmetic Science*, **33**, 338–345.

Kościński, K. (2012). Hand attractiveness—its determinants and associations with facial attractiveness. *Behavioral Ecology*, **23**, 334–342.

Miller, L., & Cox, C. (1982). For appearance's sake: Public self-consciousness and makeup use. *Personality and Social Psychology Bulletin*, **8**, 748–751.

Morrison, E. R., Morris, P. H., & Bard, K. A. (2013). The stability of facial attractiveness: Is it what you've got or what you do with it? *Journal of Nonverbal Behavior*, **37**, 59–67.

Mulhern, R., Fieldman, G., Hussey, T., Lévêque, J. L., & Pineau, P. (2003). Do cosmetics enhance female Caucasian facial attractiveness? *International Journal of Cosmetic Science*, **25**, 199–205.

Nash, R., Fieldman, G., Hussey, T., Lévêque, J. L., & Pineau, P. (2006). Cosmetics: They influence more than Caucasian female facial attractiveness. *Journal of Applied Social Psychology*, **36**, 493–504.

O'Toole, A. J., Price, T., Vetter, T., Bartlett, J. C., & Blanz, V. (1999). 3D shape and 2D surface textures of human faces: The role of "averages" in attractiveness and age. *Image and Vision Computing*, **18**, 9–19.

Porcheron, A., Mauger, E., & Russell, R. (2013). Aspects of facial contrast decrease with age and are cues for age perception. *PLoS ONE*, **8**(3): e57985.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, **57**, 199–226.

Rossi, E., Prlic, A., & Hoffman, R. (2007). A study of the European cosmetics industry. Copy available at http://edz.bib.uni-mannheim.de/daten/edz-h/gdb/07/study_eu_cosmetics_industry.pdf

Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, **38**, 1211–1219.

Russell, R. (2010). Why cosmetics work. In R. B. Adams, A. Ambady, K. Nakayama, & S. Shimojo (Eds.), *The science of social vision* (pp. 186–204). Oxford: Oxford University Press.

Said, C. P., & Todorov, A. (2011). A statistical model of facial attractiveness. *Psychological Science*, **22**, 1183–1190.

Samson, N., Fink, B., & Matts, P. J. (2010). Visible skin condition and perception of human facial appearance. *International Journal of Cosmetic Science*, **32**, 167–184.

Stephen, I. D., & McKeegan, A. M. (2010). Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception*, **39**, 1104–1110.

Swami, V., Tovée, M., & Harris, A. S. (2013). An examination of ethnic differences in actual–ideal weight discrepancy and its correlates in a sample of Malaysian women. *International Journal of Culture and Mental Health*, **6**, 96–107.

Tracy, J. L., & Beall, A. T. (2011). Happy guys finish last: The impact of emotion expressions on sexual attraction. *Emotion*, **11**, 1379–1387.

Ueno, A., Ito, A., Kawasaki, I., Kawachi, Y., Yoshida, K., Murakami, Y., … & Fujii, T. (2014). Neural activity associated with enhanced facial attractiveness by cosmetics use. *Neuroscience Letters*, **566**, 142–146.