

“Just another pretty face”: A multidimensional scaling approach to face attractiveness and variability

TIMOTHY POTTER, OLIVIER CORNEILLE, AND KIRSTEN I. RUYS
Université Catholique de Louvain, Louvain-la-Neuve, Belgium

AND

GILLIAN RHODES
University of Western Australia, Perth, Australia

Findings on both attractiveness and memory for faces suggest that people should perceive more similarity among attractive than among unattractive faces. A multidimensional scaling approach was used to test this hypothesis in two studies. In Study 1, we derived a psychological face space from similarity ratings of attractive and unattractive Caucasian female faces. In Study 2, we derived a face space for attractive and unattractive male faces of Caucasians and non-Caucasians. Both studies confirm that attractive faces are indeed more tightly clustered than unattractive faces in people's psychological face spaces. These studies provide direct and original support for theoretical assumptions previously made in the face space and face memory literatures.

A widely discussed hypothesis in the face literature suggests that attractiveness relates positively to averageness (for recent reviews, see Rhodes, 2006; Rhodes & Zebrowitz, 2002). A recent meta-analytic review has also shown that averageness and high levels of sexual dimorphism are attractive in both male and female faces (Rhodes, 2006; see also Perrett, May, & Yoshikawa, 1994). Both findings suggest that attractive faces should be perceived to be more alike than unattractive faces, because attractive faces must conform either to the population average or to an optimal direction of deviation from it. Unattractive faces, in contrast, may deviate from the population average in a number of ways.

The hypothesis that people perceive attractive faces to be more alike than unattractive faces may have important consequences for face memory, and it seems consistent with effects reported in the face memory literature. Specifically, attractive faces elicit more false memory than do unattractive faces (Corneille, Monin, & Pleyers, 2005; Monin, 2003). If attractive faces are more alike, they should be more densely clustered in face space, and so should have a higher probability of being mistaken for a previously seen face already encoded in memory (see also Lewis, 2004; Light, Kayra-Stuart, & Hollander, 1979; Mueller, Heesacker, & Ross, 1984; Vokey & Read, 1992). Not only are false alarm rates higher for attractive than for unattractive faces, but in addition hit rates are higher for unattractive faces. The latter finding would be explained if unattractive faces are less alike, and therefore can be

encoded in a more precise and discriminating fashion (see also Light, Hollander, & Kayra-Stuart, 1981; Vokey & Read, 1992).

Surprisingly enough, however, no study to date has directly examined whether attractive faces are indeed more clustered in face space. We set out to do so here, using multidimensional scaling (MDS) to derive psychological face spaces for sets that contain attractive and unattractive female Caucasian faces (Study 1) and attractive and unattractive male faces of various ethnicities (Study 2). We tested whether attractive faces are perceived to be more alike than unattractive faces by comparing the variability of attractive and unattractive faces in these derived face spaces.

STUDY 1

Method

Participants. Fifty-nine psychology undergraduates from the Université Catholique de Louvain (54 females, 5 males) participated in exchange for partial course credit.

Materials. We downloaded portraits of 80 young Caucasian females from a casting database (www.interfaces.nl). The pictures were all in black and white, and the faces had similar facial expressions. Some pictures had different poses, and we mirror-reversed some of these to ensure that each pose had a similar orientation and that the right shoulder was in the foreground. Twenty voluntary participants contacted on campus (half males, half females) rated the faces for attractiveness on a scale from 1 (*very unattractive*) to 9 (*very attractive*). On the basis of these ratings (Cronbach's $\alpha = .96$), we selected two sets of faces: 15 attractive female faces ($M = 6.9$, $SD = 0.4$, range = 6.2–7.6) and 15 unattractive female faces ($M =$

T. Potter, tim.potter@psp.ucl.ac.be

2.6, $SD = 0.4$, range = 1.9–3.3). Attractive and unattractive faces did not differ significantly in age (respectively, $M = 22.3$ years, $SD = 4.0$, and $M = 25.2$ years, $SD = 6.8$) [$t(28) = -1.4, p = .174$], and Levene’s test also showed equality of variances ($p = .34$). There was minimal pose variation between the pictures selected. If anything, poses varied slightly more for the attractive faces (because of two pictures that were framed relatively more closely than the other faces, and one face that was not as well framed as all the others), which ran counter to our hypothesis.

Procedure. The task was presented on a computer (PC compatible). All instructions were presented onscreen. Participants were asked to make similarity judgments between pairs of faces on a Likert scale from 1 (*extremely dissimilar*) to 7 (*extremely similar*). We specifically instructed participants to concentrate on the faces. We told them, “You are going to see a pair of faces in each new screen. You will have to, for each pair of faces, give a global and spontaneous judgment. There is no wrong or right answer. Simply look at the traits of these faces, and spontaneously judge their global degree of similarity.” Before proceeding to the similarity judgment task, participants were presented with the 30 faces, displayed successively in a random order at the center of the screen for 1,000 msec apiece, so that they could see the range of variation in the set. The participants then rated the similarity of 435 pairs of faces (each of the 30 faces presented once with each of the other 29 faces). One of the faces was positioned on the upper left side of the screen and the other on the upper right side. Text reminding the participants of the scale was positioned in the bottom center of the screen. The pairs of faces remained visible until participants responded by pressing a key on the keyboard (from 1 to 7). A rest screen appeared for 30 sec every 100 presentations.

Results

We first converted the ratings of similarity into dissimilarity rating matrixes for all participants. We used the INDSCAL technique (a form of weighted MDS) to process these data, which enabled us to compute the model best representing the one used by the group of participants as a whole. This technique is to be favored (Martens & Zacharov, 2000) because it accounts for differences in the importance assigned by each participant to each dimension (the *weight*). The ALSCAL procedure, which INDSCAL is based on, is limited to a maximum of six dimensions, and these were adequate for the solution to have a good fit ($R^2 = .62$, stress = 0.13) (see Figure 1, upper line).

We then extracted the coordinates for each of the 40 faces on each of the six dimensions. We computed the sum of the Euclidean distances between each face and all other faces in the same cluster (attractive vs. unattractive). For example, to compute the distance between face i and face j , we squared the deviation between the two faces on each dimension, summed these squared deviations, and took the square root of this sum. For each face, we then averaged the distances between that face and all other faces in the cluster and ran a t test. We found a highly significant effect of attractiveness [$t(28) = -5.02, p < .001$], with attractive faces being more tightly clustered ($M = 3.01, SD = 0.55$) than unattractive faces ($M = 3.38, SD = 0.48$) (see Figure 2). We also wanted to make sure our results were not due to the presence of a few highly atypical outliers in the unattractive cluster. We ran a Kolmogorov–Smirnov test, which showed the distributions of all individual Euclidean distances between faces in a given cluster to be normal (attractive faces, $p = .141$;

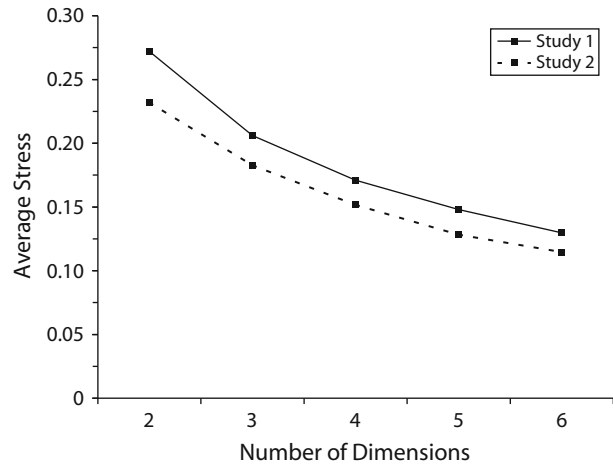


Figure 1. Averaged individual-participant stress values plotted as a function of number of dimensions for Studies 1 and 2.

unattractive faces, $p = .402$), and a Levene’s test, which showed equality of variances between these distributions ($p = .987$). Thus, data points were distributed in a balanced way around the mean. These analyses suggest that the larger mean distances of the unattractive faces are not due to a few highly atypical outliers.

Finally, we examined the relationship between the attractiveness ratings (bimodally distributed) and the (non-normally distributed) distance from each face to the cen-

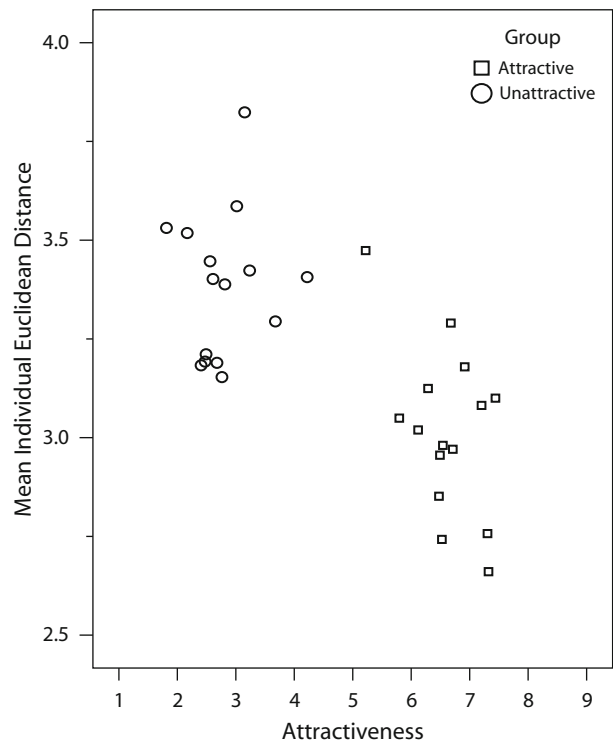


Figure 2. Scatterplot of mean individual Euclidean distances between a face and all other faces in a given group for Study 1.

troid of all faces. A nonparametric Wilcoxon rank test indicated a significant relationship between these factors ($z = -4.782, p < .001$), with attractiveness decreasing as distance to the centroid of all faces increased (see Figure 3).

STUDY 2

Method

Participants. Eighteen Caucasian participants (9 males, 9 females) were contacted on the campus of the Université Catholique de Louvain and agreed to take part in the second experiment in exchange for remuneration.

Materials. Faces were obtained from the materials of Ruys, Dijksterhuis, and Corneille (in press). Specifically, we selected 40 pictures of male faces that had been rated for attractiveness on a Likert scale from 1 to 7 (1 = *unattractive*, 7 = *attractive*): 10 pictures of attractive Caucasians ($M = 4.7, SD = 0.3, \text{range} = 4.1\text{--}5.1$), 10 of unattractive Caucasians ($M = 1.7, SD = 0.3, \text{range} = 1.4\text{--}2.2$), 10 of attractive non-Caucasians ($M = 4.7, SD = 0.5, \text{range} = 3.6\text{--}5.0$), and 10 of unattractive non-Caucasians ($M = 1.8, SD = 0.3, \text{range} = 1.4\text{--}2.2$). The non-Caucasian faces were heterogeneous in race (including African, North-African, Asian, and Indian faces). No information on age was available.

Procedure. The procedure was the same as the one used in Study 1, except that there were 780 pairs of faces.

Results

We used the INDSCAL technique, as in Study 1. The maximum number of dimensions supported by INDSCAL is six, which yielded a very satisfactory R^2 of .78 and a stress of 0.11 (see Figure 1, lower line).

We computed the average Euclidean distance between faces in each cluster (attractive Caucasian, unattractive Caucasian, attractive non-Caucasian, and unattractive non-Caucasian). The results are shown in Figure 4. We ran an ANOVA with distance as a dependent variable and race and attractiveness as independent variables. We found a significant effect of attractiveness [$F(1,36) = 69.96, p < .001$], with attractive faces being more tightly clustered ($M = 2.28, SD = 0.43$) than unattractive faces ($M = 3.19, SD = 0.40$). We also found a significant effect of race [$F(1,36) = 13.61, p < .001$], with Caucasian faces being more tightly clustered ($M = 2.54, SD = 0.66$) than non-Caucasian faces ($M = 2.94, SD = 0.51$). Finally, we found an interaction between attractiveness and race [$F(1,36) = 5.83, p < .021$], with attractive faces always being less variable than unattractive faces, but more so for Caucasian [attractive, $M = 1.95, SD = 0.24$; unattractive, $M = 3.12, SD = 0.32; t(18) = -9.30, p < .001$] than for non-Caucasian (attractive, $M = 2.62, SD = 0.31$; unattractive, $M = 3.26, SD = 0.47; p < .002$) faces.¹ As in Study 1, a Kolmogorov–Smirnov test showed the distributions of all individual Euclidean distances between faces in a given cluster to be normal (attractive faces, $p < .122$; unattractive faces, $p < .244$), and a Levene’s test showed equality of variances between these distributions ($p < .201$). These analyses suggest that the larger mean distances of the unattractive faces are not due to a few highly atypical outliers.

Thus, Study 2 generalizes attractiveness effects to faces of different genders and races. The significant effect of race

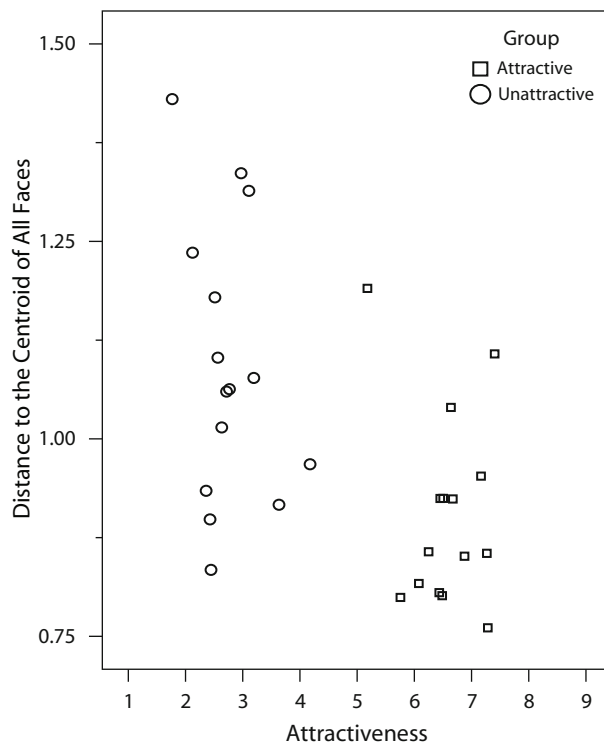


Figure 3. Scatterplot of attractiveness against distance to the centroid of all faces in Study 1.

was not surprising, considering that the non-Caucasians were heterogeneous in race. On the other hand, we can only speculate about the reasons for a race \times attractiveness interaction. Importantly, however, the attractiveness effect was obtained here on both Caucasian and non-Caucasian faces. It is also interesting to note that considering their ethnic diversity, attractive non-Caucasian faces were nonetheless more tightly clustered than unattractive Caucasian faces were. The attractiveness effect was thus sufficiently strong to make attractive faces of diverse races more similar than unattractive Caucasian faces.

As in Study 1, we examined the relationship between attractiveness (bimodally distributed) and the distance from each face to the centroid of all faces. These results are shown in Figure 5. Unlike in Study 1, here distance to the centroid *was* normally distributed, so we ran an ANOVA with distance to the centroid as a dependent variable and attractiveness and race as independent variables. We found that attractive faces were closer to the centroid ($M = 0.80, SD = 0.16$) than were unattractive faces ($M = 1.19, SD = 0.21$) [$F(1,36) = 42.57, p < .001$]. No other effects were significant (all other p values $> .145$).²

GENERAL DISCUSSION

In line with the theoretical assumptions made in both the face space and face memory literatures, attractive faces were more tightly clustered than unattractive faces in the psychological face spaces derived from similarity ratings of the faces. In Study 1, the female faces were obtained

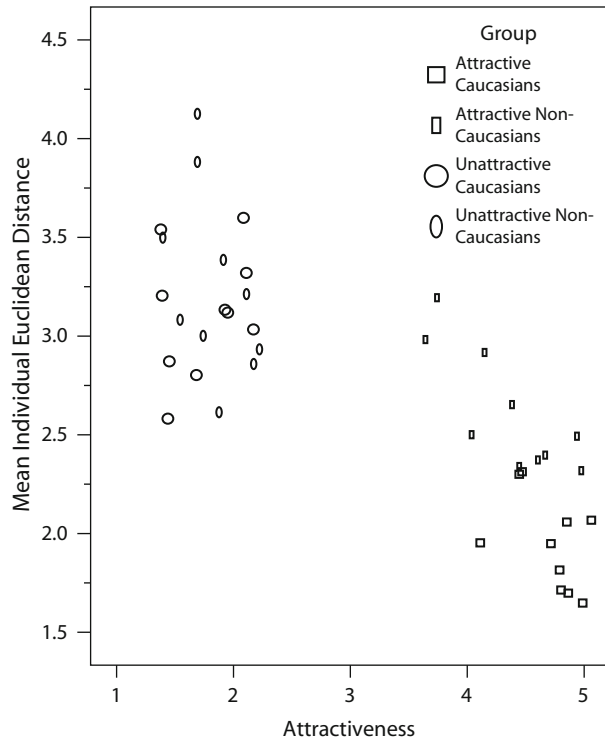


Figure 4. Scatterplot of mean individual Euclidean distances between a face and all other faces in a given group for Study 2.

from a casting database, and may therefore have been slightly biased toward relatively attractive faces. Nevertheless, a clear difference in variability was observed for the attractive and unattractive sets. Furthermore, the same difference was replicated in male faces, both Caucasian and non-Caucasian, which were obtained from a different source (Ruys et al., in press). Therefore, the present research provides original and direct evidence for the greater variability of unattractive faces. This finding appears to be a robust effect, since in the present two studies it generalized to faces from various gender and ethnic identities. This finding seems consistent with the casual observation that beauty transcends racial divides, with models looking relatively alike whatever their ethnicity. It also offers a clear explanation of attractiveness effects in memory for faces (Corneille et al., 2005; Light et al., 1979; Monin, 2003; Mueller et al., 1984; Vokey & Read, 1992): The clustering of attractive faces in face space explains the high false alarm rates observed for novel attractive faces, which are likely to produce false matches with familiar faces in face space. Higher hit rates for unattractive than for attractive faces are also explained, because the former have more unique representations.

It seems interesting to relate the present work to perceptual fluency effects examined in the categorization and face memory literatures. Because attractive faces are less distinctive (i.e., more tightly clustered) in the face space, they are processed more fluently. In turn, people mistake this enhanced fluency of attractive faces for face familiarity, contributing to increased false memory for these

faces (see Monin, 2003). Corneille et al. (2005) reported direct evidence of fluency effects in the false memory for attractive faces (for a related point, see also Kleider & Goldinger, 2004). It should be noted that recent research has also reported evidence for a direct impact of fluency on similarity judgments, with stimuli perceived to be more similar when they are rendered more perceptually fluent (Blok & Markman, 2005). The relatively better fluency of attractive faces may have contributed to increasing the similarity ratings for these faces. However, our claim is that attractive faces may be processed more fluently because they are less distinctive in the face space.

Finally, it is worth discussing how the present findings relate to the prior face space literature, which has assumed that averageness is correlated with density (Valentine, 1991). In the present two studies, we obtained support for this assumption, with face attractiveness decreasing with distance from the overall centroid (the average of all faces in the model). Only two studies in the face space literature have relied on an MDS procedure to examine the assumption that attractive faces are located closer to the face space centroid. These studies (Busey, 1998; Lee, Byatt, & Rhodes, 2000) provided mixed findings and did not directly control for the attractiveness of the faces. Rather, they relied on the assumption that composite faces should be on average more attractive than the faces used to compose them. However, it is important to note that the MDS technique does not constrain variability to decrease with distance from the MDS centroid. Thus, our finding that attractiveness is negatively associated with both variability

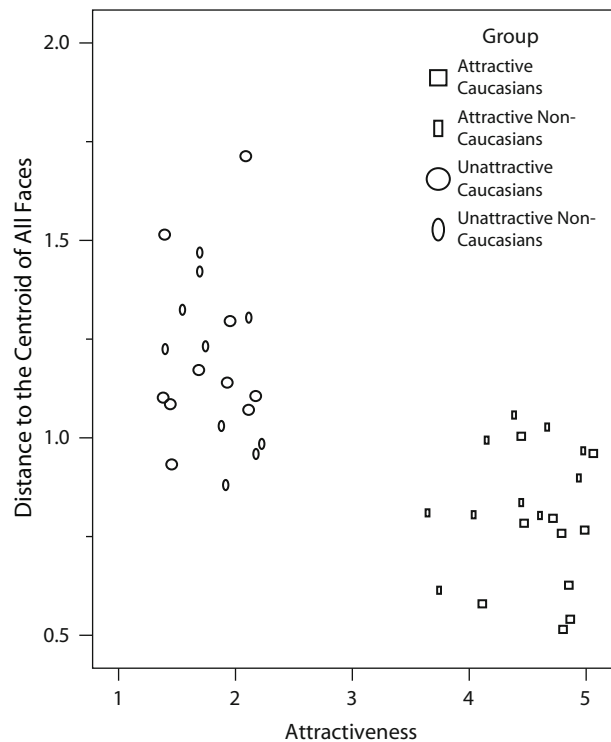


Figure 5. Scatterplot of attractiveness against distance to the centroid of all faces in Study 2.

and distance from the face space centroid is not a statistical artifact. Rather, it supports a theoretical assumption of the face space framework. We note, however, that this conclusion assumes that a centroid calculated from samples of attractive and unattractive faces is close to the average for the entire face space.

The face space model and the MDS conceptualization it entails represent a convenient theoretical tool for understanding a variety of face-processing effects. Ironically, however, it is difficult to use MDS to directly examine some specific predictions derived from this model. This is because one would need a random sample of a significant number of faces in order to locate relatively accurately a participant's face space centroid. Unfortunately, the number of similarity ratings required becomes exponentially large as additional faces are considered for pairwise similarity judgments. By focusing on face density, we were able to bypass this limitation yet obtain findings that are fully consistent with theoretical claims made in both the face space and face memory literatures.

Conclusion

By relying on an MDS approach, we were able to provide original evidence that attractive faces are more densely clustered than unattractive faces in psychological face spaces. This finding is consistent with assumptions made in the face space and face memory literatures. By focusing on the face variability hypothesis, we were able to bypass the issue of determining participants' "real" face space centroids. That attractive faces are more tightly clustered than unattractive ones provides a clear account of why an attractive face tends to be seen as just *another* pretty face.

AUTHOR NOTE

This research benefited from FSR Grant PSORC505 awarded to O.C. Correspondence concerning this article may be addressed to T. Potter, Université Catholique de Louvain, PSP-PSOR, 10, Place du Cardinal Mercier, B-1348 Louvain-la-Neuve, Belgium (e-mail: tim.potter@psp.ucl.ac.be).

REFERENCES

- BLOK, S. V., & MARKMAN, A. B. (2005). Fluency in similarity judgments. In B. G. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 280-285). Mahwah, NJ: Erlbaum.
- BUSEY, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science*, *9*, 476-483.
- CORNEILLE, O., MONIN, B., & PLEYERS, G. (2005). Is positivity a cue or a response option? Warm glow vs. evaluative matching in the familiarity for attractive and not-so-attractive faces. *Journal of Experimental Social Psychology*, *41*, 431-437.
- KLEIDER, H. M., & GOLDINGER, S. D. (2004). Illusions of face memory: Clarity breeds familiarity. *Journal of Memory & Language*, *50*, 196-211.
- LEE, K., BYATT, G., & RHODES, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, *11*, 379-385.
- LEWIS, M. B. (2004). Face-space-R: Towards a unified account of face recognition. *Visual Cognition*, *11*, 29-69.
- LIGHT, L. L., HOLLANDER, S., & KAYRA-STUART, F. (1981). Why attractive people are harder to remember. *Personality & Social Psychology Bulletin*, *7*, 269-276.
- LIGHT, L. L., KAYRA-STUART, F., & HOLLANDER, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 212-228.
- MARTENS, W. L., & ZACHAROV, N. (2000, September). *Multidimensional perceptual unfolding of spatially processed speech I: Deriving stimulus space using INDSCAL*. Paper presented at the 109th Convention of the Audio Engineering Society, Los Angeles.
- MONIN, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality & Social Psychology*, *85*, 1035-1048.
- MUELLER, J. H., HEESACKER, M., & ROSS, M. J. (1984). Likability of targets and distractors in facial recognition. *American Journal of Psychology*, *97*, 235-247.
- PERRETT, D. I., MAY, K. A., & YOSHIKAWA, S. (1994). Facial shape and judgements of female attractiveness. *Nature*, *368*, 239-242.
- RHODES, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199-226.
- RHODES, G., & ZEBROWITZ, L. A. (Eds.) (2002). *Facial attractiveness: Evolutionary, cognitive, and social perspectives*. Westport, CT: Ablex.
- RUYS, K. I., DIJKSTERHUIS, A., & CORNEILLE, O. (in press). On the (mis)categorization of unattractive brides and attractive prostitutes: Extending evaluative congruency effects to social category activation. *Experimental Psychology*.
- VALENTINE, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology*, *43A*, 161-204.
- VOKEY, J. R., & READ, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, *20*, 291-302.
- YZERBYT, V. Y., MULLER, D., & JUDD, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, *40*, 424-431.

NOTES

1. We did not control for age as a factor in Study 2. Since information on age was unavailable, we decided to run a posttest in which we asked 20 undergraduate students (half male, half female) to rate the ages of the pictures presented. We then ran an ANOVA with rated age as a dependent variable and attractiveness and race as independent variables. In this model, attractiveness came out as a significant factor and race as a marginally significant one, thereby possibly confounding attractiveness with age and (marginally) with race. This situation may have been problematic, so we decided to control for perceived age by regressing the Euclidean distances on attractiveness, race, attractiveness \times race, age \times attractiveness, age \times race, and age \times attractiveness \times race, after centering each of these factors (see Yzerbyt, Muller, & Judd, 2004). All effects remained significant [attractiveness, $t(32) = -7.27$, $p < .001$, $\beta = -.81$; race, $t(32) = -3.80$, $p < .001$, $\beta = -.42$; attractiveness \times race, $t(32) = -2.13$, $p < .041$, $\beta = -.24$]. We found neither main nor moderating effects of perceived age in this model (all p values $> .30$).

2. For the distance to the centroid, we controlled perceived age in a fashion similar to the one used for Euclidean distance. Consistent with predictions, attractiveness was significant in this model [$t(32) = -4.99$, $p < .001$, $\beta = -.67$]. We also found a marginally significant interaction between attractiveness and race [$t(32) = -1.98$, $p < .057$, $\beta = -.27$]. None of the other effects were significant.

(Manuscript received February 7, 2006;
revision accepted for publication April 28, 2006.)