Supplementary Materials
Similarity Ratings

(1) Ruling out category priming

Faces of different races represent distinct categories and might be coded in different locations in face space (e.g., Jaquet et al., 2008). This raises the question whether adaptation with an antiface of the same race as the target face (which is true for the same-race trajectory in our experiment) gives a recognition advantage over adaptation to an antiface of a different race (as on the mixed-race trajectory). In this case, better performance on the same-race trajectory that we found in this study might just be due to priming of the correct face category on this but not on the mixed-race trajectory. To rule out such an account, we introduced the non-central trajectory, where the antiface is of the same race as the target faces (as on the same-race trajectory), but is created relative to an individual identity that is taken from a non-central location in face space. To make sure that the similarity between the targets and the reference face are the same on both trajectories (and that aftereffects on these trajectories are thus comparable), we chose these non-central reference faces on the basis of similarity ratings. The perceptual distance between the target faces and the same-race average was thus matched to the perceptual distance between the targets and the two non-central reference faces that were used in the main study.

Method

Participants

Fourteen young Caucasian adults from the student pool of the University of Western Australia volunteered to do the rating experiment. Eleven participants rated Caucasian and nine participants Asian faces (with six students rating only Caucasian target faces, four only Asian target faces, and five students doing both versions of the experiment, with a time lag of at least 6 days between sessions). None participated in the main experiment.
Stimuli

Thirteen Asian and thirteen Caucasian faces from the pool that was used to create the average faces (twenty Asian and twenty Caucasian male faces) were used for this rating experiment.

Procedure

The task was run on a 20-inch LCD screen iMac OS X, version 10.5.6. Stimuli were presented using the presentation software Superlab 4.0.6, at a screen resolution of 1680 by 1050 pixels. Participants were first presented with instructions on the computer screen explaining what would be required of them.

Thirteen faces from each set of twenty faces used to create the Asian and Caucasian averages were paired with each of the four target faces of the same race set (Asian faces with Asian targets and Caucasians with Caucasians, respectively). All faces, including the target faces, were also paired with their own-race average. Participants were presented with one face pair at a time, along with a rating scale, ranging from 1 (not at all similar) to 7 (very similar). Every face pair (69 per race) was shown twice, once in each left-right arrangement. Participants were instructed to answer as quickly as possible and to use the full range of the scale, using the number keys on the computer keyboard. To expose them to the range and appearance of the face stimuli used in the rating, participants were shown 22 other single faces at the beginning of the experiment. They had no answer to give and these faces were not used again later for the rating. The whole experiment took about 10 minutes to complete.

Results

The perceptual similarity ratings for each face pair were averaged across participants, for each race separately. The average distance between the Asian targets and the Asian average face was 3.45 (SEM = 0.26). The face whose ratings came closest to that had an average distance to the targets of 3.53 (SEM = 0.36). The average distance between the Caucasian targets and the
Caucasian average face was 3.92 (\(SEM = 0.24\)). The face whose ratings came closest to that had an average distance to the targets of 3.66 (\(SEM = 0.40\)). These two faces with an average distance to the targets that was very close to the one between the targets and their own-race average face were thus used in lieu of an average face to create the ‘non-central’ morph trajectory in the adaptation experiment. Pair-wise t-tests show that the similarity ratings for targets and average and the ratings for targets and these two non-central faces did not differ significantly [Asian faces: \(t(8) = -0.25, p = 0.809 (SEM = 0.320)\); Caucasian faces: \(t(10) = 1.09, p = 0.312 (SEM = 0.244)\)].

(2) Length of trajectories and extremity of adaptors

In the main experiment, we found that aftereffects were larger for same-race than mixed-race trajectories. Since aftereffects are larger for opposite than non-opposite adapt-test pairs, we argued that this suggests that identity is coded using a same-sex average. However, larger aftereffects on same-race trajectories could also be explained by two alternative accounts. First: The target faces are likely to be more similar to an average face of the same race than to a mixed-race average. If so, the larger aftereffects that we found for (shorter) same-race trajectories might actually represent smaller absolute shifts in perceptual distance than the smaller aftereffects observed for (possibly longer) generic trajectories.

Second: Since the antifaces are also likely to be more similar to a same-race than a mixed-race average, adaptors might be less extreme on same-race than mixed-race trajectories. While face aftereffects are generally larger for more extreme adaptors (Robbins et al., 2007), we found the opposite pattern: larger aftereffects on same-race trajectories. Hence, the expected difference in adaptor extremity would be in the wrong direction to explain our results.

To assess the length of the three trajectories and to estimate adaptor-test similarity for each trajectory, we obtained ratings of similarity of the target faces to the averages (80% vs 0% pairs) and of the adapting antifaces to the averages (-50% vs 0% pairs).
Method

Participants

Twelve young Caucasian adults were recruited from the student pool at the University of Western Australia. None had participated in the main experiment.

Stimuli

The same face stimuli as in the main experiment were used. For the present rating study, stimuli consisted of the four target faces (for each race, respectively) morphed with each of the three reference faces, at an identity strength of 80%. For each of the resulting faces (12 faces in each race), the antiface (-50%) was included, as well as the three reference faces for each race. There were thus 27 faces for each race.

Procedure

The task was run on a 20-inch LCD screen iMac OS X, version 10.5.6. Stimuli were presented using the presentation software Superlab 4.0.6, at a screen resolution of 1680 by 1050 pixels. Participants were first presented with instructions on the computer screen explaining what would be required of them.

For each trajectory, two comparisons were made: target face (80%) vs. reference (0%), and reference (0%) vs. antiface (-50%). Pairs of faces were shown sequentially, and each comparison was presented twice with sequence order presentation alternated. In total, there were 96 trials (3 trajectories x 4 target faces x 2 races, x 2 comparisons, x 2 sequence orders), presented in two blocks of 48 trials of only one race, but with trajectories presented intermixed (randomized). Block presentation order was counterbalanced across participants.

Each block consisted of an initial viewing phase followed by a rating phase. In the viewing phase, participants were shown all pair combinations and were instructed to look closely at each face pair to get a general idea of how the pairs differ in similarity. Once all pair combinations were viewed, participants were instructed on the screen to look at all pairs again and this time rate them for
Similarity, using a numbered 7-point scale ranging from 1 (not at all similar) to 7 (very similar). Participants were instructed to answer as quickly as possible and to use the full range of the scale, using the number keys on the computer keyboard. The whole experiment (i.e., both blocks of Asian and Caucasian faces, respectively) took approximately 15 minutes to complete.

Results

Similarity target – reference faces (80% vs. 0%)

Mean similarity ratings were calculated for each trajectory and face-race for each participant. A two-way ANOVA with trajectory as within-subjects factor and face race as between-subjects factor was conducted on the mean similarity ratings. There was a significant effect of trajectory \( [F(2,44) = 14.500, p < 0.001, \text{partial } \eta^2 = 0.397] \). Pair-wise t-tests show that pairs from same-race trajectories (Mean = 3.61, SEM = 0.21) and pairs from non-central trajectories (Mean = 3.52, SEM = 0.22) were rated as significantly more similar than pairs from mixed-race trajectories (Mean = 2.52, SEM = 0.24; same vs. mixed: \( t(23) = 3.860, p = 0.001 \); non-central vs. mixed: \( t(23) = 0.4220, p < 0.001 \)). Pairs from non-central trajectories, however, had the same perceptual similarity as pairs from same-race trajectories \( (t(23) = 0.759, p = 0.456) \). This is in accordance with the fact that the non-central reference faces were explicitly chosen on the basis of similarity ratings, to match their distance to the targets to the distance between the targets and the same-race averages. There was no significant effect of target race \( [F(1,22) = 0.108, p = 0.746, \text{partial } \eta^2 = 0.005] \) nor an interaction with trajectory \( [F(2,44) = 0.936, p = 0.4, \text{partial } \eta^2 = 0.041] \). See Table 1 for means and SEMs.

| Reference / Trajectory | Face Race |  |
|------------------------|-----------|
|                        | same-race | mixed-race | non-central |
| Asian                  | 3.73 (0.36) | 2.40 (0.25) | 3.34 (0.38) |
| Caucasian              | 3.50 (0.23) | 2.64 (0.41) | 3.70 (0.25) |
| All                    | 3.61 (0.21) | 2.52 (0.24) | 3.52 (0.22) |

Table 1. Mean (SE) similarity of target faces to same-race and mixed-race reference faces (80% vs. 0%)
Correction for differences in test trajectory length

The similarity ratings for the target-reference (80% vs. 0%) pairs were used to estimate and adjust any difference in test trajectory lengths. Target faces were rated as 30% less similar to the generic than the same-race averages (2.52 vs. 3.61) and 3% less similar to the non-central than the same-race average (3.52 vs. 3.61). The generic-trajectory aftereffect was therefore scaled (multiplying) by 1.3, and the non-central-trajectory aftereffect was multiplied by 1.03. Then, the effect of trajectory was reexamined in a two-way ANOVA with trajectory as repeated measures factor and face race as between-subjects factor. The ANOVA revealed a significant effect of trajectory \(F(2,50) = 4.844, p = 0.012, \text{partial } \eta^2 = 0.162\], but no main effect of face race \(F(1,25) = 1.678, p = 0.207, \text{partial } \eta^2 = 0.063\] and no interaction \(F(2,50) = 0.570, p = 0.569, \text{partial } \eta^2 = 0.022\]. Pair-wise t-tests show that aftereffects remained larger for same than for mixed and non-central trajectories \([\text{same vs. mixed: } t(26) = 2.385, p = 0.003; \text{same vs. non-central: } t(26) = 3.457, p = 0.002]\], whereas the aftereffects for mixed and non-central trajectories were not significantly different from each other \([t(26) = 0.219, p = 0.829]\).

Correction for differences in extremity of adaptors

Antifaces were rated as more similar to non-central \((M = 6.21, \text{SEM} = 0.17)\) than to mixed \((M = 5.13, \text{SEM} = 0.17)\) and race-specific averages \((M = 5.33, \text{SEM} = 0.20)\). Pair-wise t-tests showed that the difference in similarity between non-central trajectories and both other trajectories was significant \([\text{same vs. non-central: } t(23) = -4.892, p < 0.001; \text{mixed vs. non-central: } t(23) = 4.308, p < 0.001]\), but not the difference on same-race and mixed-race trajectories \([t(23) = 0.866, p = 0.396]\). There was a significant main effect of trajectory \([F(2,44) = 12.697, p < 0.001, \text{partial } \eta^2 = 0.366]\), but no main effect of face race \([F(1,22) = 0.554, p = 0.465, \text{partial } \eta^2 = 0.025]\] and no interaction \([F(2,44) = 0.459, p = 0.635, \text{partial } \eta^2 = 0.020]\).

Since the non-central antifaces are more similar to the reference faces, they are less extreme adaptors than the antifaces on the other trajectories.
Smaller aftereffects for non-central antifaces could in principal be explained by these differences. We thus scaled the aftereffects again, relative to these differences. Here, contrary to the trajectory length measured above, a higher value (i.e., more similar antiface – reference pairs) is expected to lead to a weaker aftereffect. The race-specific antiface is 4% more similar to the average than the mixed-race antiface, and the non-central antiface is 17% more similar to its reference than the mixed-race antiface to the mixed-race average. We therefore multiplied the (already adjusted, see above) AEs, again, by 1.04 (same-race trajectory) and 1.17 (non-central trajectory).

Then, the effect of trajectory was re-examined in a two-way ANOVA with trajectory as repeated measures factor and face race as a between-participants factor. The ANOVA revealed a significant main effect of trajectory \([F(2,50) = 4.323, p = 0.019, \text{partial } \eta^2 = 0.147]\), but no main effect of face race \([F(1,25) = 1.443, p = 0.241, \text{partial } \eta^2 = 0.055]\) and no interaction \([F(2,50) = 0.557, p = 0.577, \text{partial } \eta^2 = 0.022]\). Pair-wise t-tests show that aftereffects remained larger for same than for mixed and non-central trajectories \([\text{same vs. mixed: } t(26) = 2.523, p = 0.018; \text{same vs. non-central: } t(26) = 2.919, p = 0.007]\), whereas the aftereffects for mixed and non-central trajectories were not significantly different from each other \([t(26) = 0.110, p = 0.913]\). See Table 2 for means and SEMs.

<table>
<thead>
<tr>
<th>Reference / Trajectory</th>
<th>Face Race</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>same-race</td>
<td>mixed-race</td>
<td>non-central</td>
</tr>
<tr>
<td>Asian</td>
<td>5.35 (0.28)</td>
<td>4.93 (0.24)</td>
<td>6.10 (0.20)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>5.31 (0.29)</td>
<td>5.32 (0.22)</td>
<td>6.31 (0.30)</td>
</tr>
<tr>
<td>All</td>
<td>5.33 (0.20)</td>
<td>5.13 (0.17)</td>
<td>6.21 (0.17)</td>
</tr>
</tbody>
</table>

Table 2. Mean (SE) similarity of adapting antifaces to same-race and mixed-race reference faces (-50% vs. 0%).