Social categories shape the neural representation of emotion: evidence from a visual face adaptation task

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Social categories shape the neural representation of emotion: evidence from a visual face adaptation task

Marte Otten[1,2]* and Mahzarin R. Banaji[1]

[1] Department of Psychology, Harvard University, Cambridge, MA, USA
[2] Department of Social Psychology, University of Amsterdam, Amsterdam, Netherlands

A number of recent behavioral studies have shown that emotional expressions are differently perceived depending on the race of a face, and that perception of race cues is influenced by emotional expressions. However, neural processes related to the perception of invariant cues that indicate the identity of a face (such as race) are often described to proceed independently of processes related to the perception of cues that can vary over time (such as emotion). Using a visual face adaptation paradigm, we tested whether these behavioral interactions between emotion and race also reflect interdependent neural representation of emotion and race. We compared visual emotion aftereffects when the adapting face and ambiguous test face differed in race or not. Emotion aftereffects were much smaller in different race (DR) trials than same race (SR) trials, indicating that the neural representation of a facial expression is significantly different depending on whether the emotional face is black or white. It thus seems that invariable cues such as race interact with variable face cues such as emotion not just at a response level, but also at the level of perception and neural representation.

Keywords: emotion, face perception, racial prejudice, visual adaptation

Faces contain a wealth of social information, conveying cues about identity, personality, intentions, and emotions. It is, therefore, not surprising that social cues encoded by faces are extracted quickly and efficiently (Mouchetant-Rostaing and Giard, 2003; Blau et al., 2007; Ito and Bartholow, 2009).

A number of behavioral studies indicate that cues that are used to establish the identity of a face, i.e., invariable cues about face structure such as race and gender, influence the processing of variable aspects of the face such as emotional expression. Especially the representation of social category information about race (an invariable facial feature) and the emotional expression of the face (a variable facial feature) appear to be closely interlinked. A number of recent studies have shown that emotional expressions are perceived differently and more effectively depending on the race of a face (Elfenbein and Ambady, 2003; Hugenberg and Bodenhausen, 2003; Bestelmeyer et al., 2010; Bijlstra et al., 2010). In addition, the perception and encoding of race cues can depend on facial emotional expression (Hugenberg and Bodenhausen, 2004; Ackerman et al., 2006; Richeson and Trawalter, 2008). Together, these studies suggest that perceptual cues related to the race and emotion expression of a face are not processed independently. Instead, an invariant feature such as race can alter the perception of variable features such as emotional expressions, and variable features such as emotional expressions can in turn change the perception of invariant features such as the race of a face.

Although there is clear behavioral evidence that invariable race cues and variable emotion cues are processed interdependently, there is very little evidence of interactive processing of emotion and race cues at a neurological level. On the contrary, influential cognitive and neuroscientific models of face perception assert that invariable cues such as race and gender are processed independently from variable facial cues such as emotional expression and gaze direction (Bruce and Young, 1986; Haxby et al., 2000). Studies that directly looked at brain processes involved in the perception of emotion and social categories show clear evidence of independent and parallel processing of the two. For example, Kubota and Ito (2007) showed subjects black and white faces with a range of emotional expressions. The Event-related Potentials (ERPs) evoked by these different faces showed that specific race- and emotion-related components both emerge early but proceed in parallel, showing no reliable interactions.

Similarly, several brain imaging studies have shown that the processing of social category information of a face engages different brain areas than regions involved in processing emotion-related information, again suggesting independence. In general, face processing consistently activates the (right or bilateral) lateral fusiform gyrus, also referred to as the fusiform face area, and the superior temporal sulcus (Kanwisher et al., 1997; Halgren et al., 1999). However, these two regions seem to be independently engaged in processing variable and invariable face cues. Invariable face cues, such as those related to the identity of the face, engage the fusiform face area. Processing of variable face cues, such as those related to emotion or eye gaze, engage the superior temporal sulcus (Hoffman and Haxby, 2000). Lesion studies show the same pattern of results (Humphreys et al., 1993). These studies thus strongly indicate that brain processes underlying the perception of race and emotion proceed independently.

Taken together, the current pattern of results is inconsistent at best. On the one hand, human behavior shows that the processing and representation of facial emotion and race are likely

interlinked. On the other hand, studies directly measuring brain processes that underlie emotion and race perception suggest independent processing of these two types of information.

There are several ways to explain such an inconsistency. One possibility is that the brain represents variable cues like emotion and invariable cues like race independently, as reflected by the imaging and ERP findings, and that subsequent behavioral interactions purely take place at the stage of response preparation, or execution. However, it is also possible that, while the initial processing of emotion- and race-related facial cues is truly independent, as is shown by the multitude of imaging studies, these processing streams interact at some point (possibly at the amygdala, see also Haxby et al., 2002) leading to final interdependent neural representations of emotion and social category as observed in behavioral studies. The key difference between these explanations is thus whether or not the neural representation of emotional expressions changes with the race of the face (and vice versa).

In two experiments, we tested whether identical emotional expressions, expressed by faces with different races (DRs) rely on the same, or on different perceptual representations in the brain. This was tested using a visual adaptation paradigm. Visual adaptation to a face with specific characteristics influences subsequent perception of faces with similar characteristics, notably by diminishing perception of the characteristics of the test face that match the adapting face. For example, after viewing an explicitly angry face, subsequent ambiguous "angry-frightened" faces appear relatively less angry and more frightened (Webster et al., 2004). These face adaptation effects seem to rely on adjustments of relatively high-level representations of different facial features and cues, since the aftereffects are not sensitive to changes in size, orientation, or position (Watson and Clifford, 2003).

Adaptation tasks such as these are specifically suitable to test whether information about social categories and emotional expressions are neurally independent or not. If emotion is represented independently of social categories such as race, then black and white faces expressing the same emotion should engage identical neural substrates, resulting in similar aftereffects for emotional expressions independent of whether the first "adapting" face and the second "test" face are from the same or different racial categories. On the other hand, if the underlying mental representations of emotional expressions differ to some extent as a function of race, then the emotion aftereffect should be diminished or absent when the adapting and test faces differ in race.

Although visual aftereffects such as the emotion aftereffect are behavioral measures, their occurrence provide direct information about the neural representations that underlie our perception. The psychophysical aspects of the aftereffect thus allows us to test whether emotional expressions are represented independently of invariable facial cues, or whether the neural representations of emotions are influenced by perceived social categories such as race.

EXPERIMENT 1

In experiment 1 we used visual aftereffects to measure the level of overlap in the neural representation of emotions expressed by black and white faces. In different blocks, the relationship between the adapting face and the test faces was varied. In the Same Race/Same Identity (SR/SI) condition, the test face and the adapting face had the same identity and thus the same race (SR) as well. In the Different Race/Different Identity (DR/DI) condition, adapting faces had a DR, and thus a different identity than the test faces. DR/DI and SR/SI trials do not only differ in the racial (dis)similarity between the adapting and test faces, but also in the absence or presence of identity differences between the adapting and test face. This introduces additional differences between the adapting and test face than just racial features, which could account for any decrease or even absence of emotion adaptation (Wu et al., 2009). To test how much of the difference in emotion aftereffects between DR/DI and SR/SI trials can be attributed to these differences in identity and background facial features (and thus not to differences in race) we added a Same Race/Different Identity (SR/DI) condition. In the SR/DI trials, the adapting and test faces were of the SR, but had different identities. Therefore, if any differential effects observed in the DR/DI condition are due to differences in identity and/or background features, these same differential effects should also be present in the SR/DI condition. If the decrease or absence of adaptation in the DR/DI condition is due to differences in race, these effects should not be present in the SR/DI condition. **Figure 1** shows example trials for each of the three conditions.

METHODS

Participants

Twenty-two participants participated in Experiment 1. The average age was 20 years, and 16 of them were women. Of the 22 participants, two were African-American, nine were European-American, one was Hispanic, one was Middle Eastern, and nine were Asian.

Stimuli

The faces that functioned as adapting and test stimuli were created using Facegen Software. All stimuli were based on four different identities, two white males and two black males. For each identity a 100% angry and a 100% frightened face was generated to function as the adapting stimulus. In addition 13 test faces were generated. These test faces had an ambiguous expression, ranging from 20% angry/80% afraid to 80% angry/20% afraid in 5% increments. When presented on the screen, the faces had a height of approximately 10 cm, and a width of about 5.5 cm.

Procedure

At the start of the experiment, each participant was randomly assigned one black and one white identity which would function as the identity of the test faces throughout the experiment.

After subjects had read and signed the informed consent, the experiment started with a practice block, in which subjects viewed and categorized just the ambiguous test faces. The practice block consisted of 104 trials, with four repeats for each of the 26 different test faces that were used for that participant. Each face was centrally presented for 300 ms, and was followed by a response screen that consisted of a centrally presented question mark, and the words AFRAID and ANGRY at the bottom of the screen, at the right and left hand corner. Subjects responded using the "a" key (left hand button press for a frightened face) or the "l" key (right hand button press for an angry face).
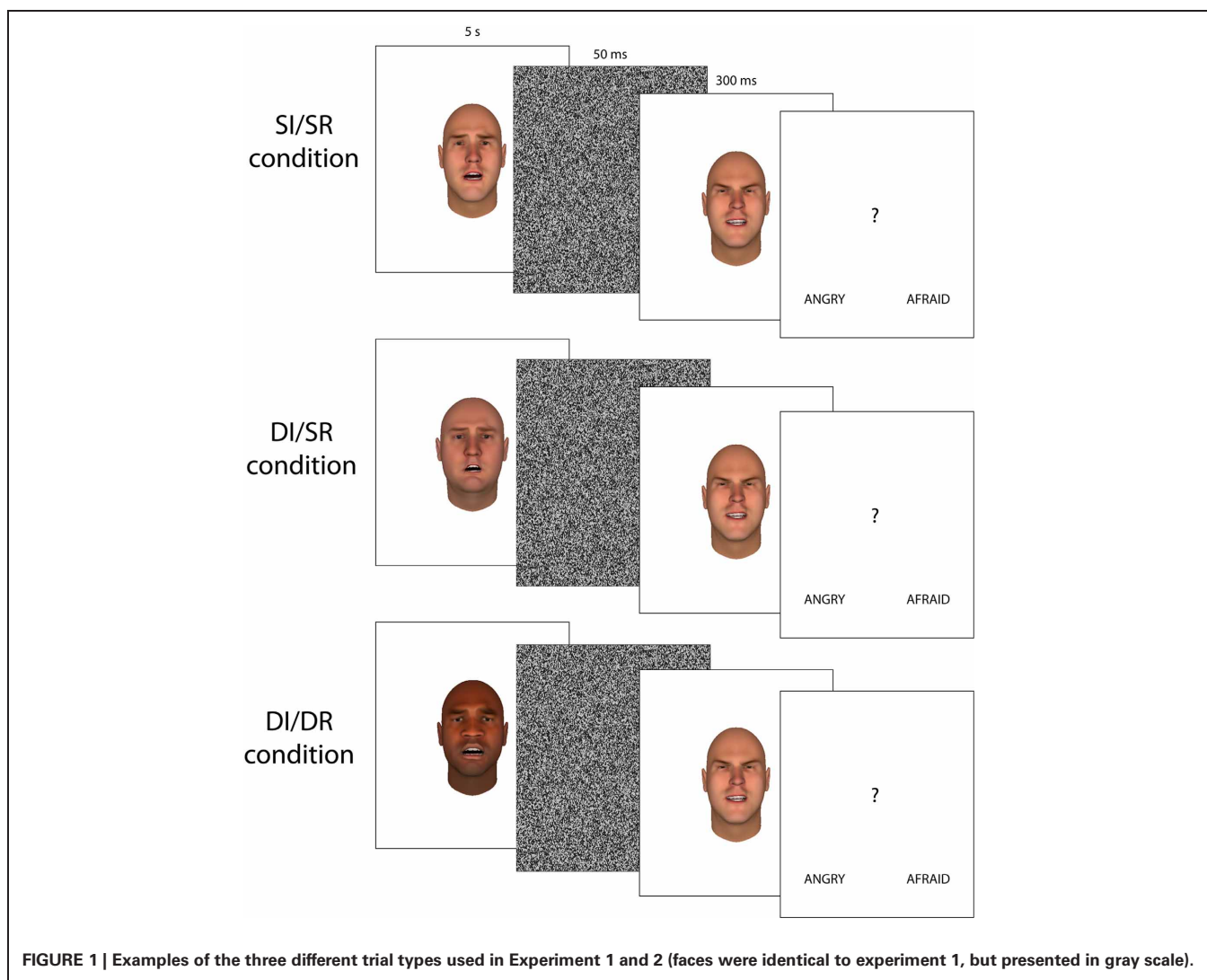
FIGURE 1 | Examples of the three different trial types used in Experiment 1 and 2 (faces were identical to experiment 1, but presented in gray scale).

Adaptation trials consisted of a 5 s adaptation period in which subjects viewed an angry or a frightened face. The adapting face was followed by a 50 ms pattern mask, after which an ambiguous test face was presented for 300 ms. Participants were then asked to indicate whether this ambiguous test face had an angry or a frightened expression with a left or right-hand button press using the same response screen as in the practice trials.

The relationship between the adapting and test faces differed over blocks. In the SR/SI block, the adapting face had the same identity as the test face. In the SR/DI block the adapting and test face were always different people, but with the SR. Therefore, in both the SR/SI and the SR/DI block, a black adapting face was always followed by a black test face, and a white adapting face was always followed by a white test face. In the DR/DI block the adapting and test face were always of a DR, and, therefore, also of a different identity, such that a black adapting face was always followed by a white test face and a white adapting face was always followed by a black test face.

Subjects completed three blocks, each consisting of 52 trials [13 ambiguous faces, two identities (1 black and 1 white test identity) and two emotions (angry and afraid adapting face)]. The order in which the three experimental blocks were presented was randomized over subjects.

Analysis
For each of the three conditions, SI/SR, DI/SR, and DI/DR, the percentage of "frightened" responses was calculated for the two adapting emotions. The difference between the proportion of "frightened" responses following angry vs. frightened adapting faces was taken to indicate the magnitude of the emotion aftereffect. An ANOVA with the factors Adapting Emotion (angry vs. frightened) and Adaptation Condition (SI/SR, DI/SR, and DI/DR) tested for differences in the magnitude of the emotion aftereffect between the three adaptation conditions. In case of a significant interaction between Adapting Emotion and Adaptation Condition *Post-hoc* *t*-tests were performed to test the nature that interaction. *F*-tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse–Geisser or Huynh–Feldt correction where appropriate. Uncorrected degrees of freedom and corrected *p*-values are reported.

RESULTS

After seeing an angry adapting face, subjects judged subsequent ambiguous test faces to be more afraid, whereas adaptation to a frightened face made subsequent ambiguous faces seem more angry [$F(1,21) = 260.7$, $p < 0.001$, partial $\eta^2 = 0.93$]. Crucially, as **Figure 2** shows, this emotion aftereffect was moderated by the relationship between the adapting and ambiguous test face [$F(2,42) = 4.4$, $p = 0.02$, partial $\eta^2 = 0.18$]. *Post-hoc t*-tests showed that the emotion aftereffect was smaller in DI/DR trials compared to both the SI/SR [$t(21) = 3.1$, $p = 0.005$, Cohen's $d = 0.96$] and DI/SR trials[$t(21) = 2.2$, $p = 0.04$, Cohen's $d = 0.67$]. The emotion aftereffect was comparable in size for SI/SR and DI/SR trials [$t(21) = 0.6$, $p = 0.58$, Cohen's $d = 0.17$]. Thus, a switch in race between the adapting and the test face significantly reduces the emotion aftereffect. In addition, this reduction in the size of the aftereffect cannot be attributed to the simple difference in identity (which accompanies the difference in race) between the adapting and test face, since a difference in identity alone does not lead to a similar decrease of the aftereffect.

EXPERIMENT 2

The faces in Experiment 1 were presented in color, and the color transition within DR trials (white adapting to black test face, or black adapting to white test face) was obviously larger than within SR trials. It could thus be that the reduction of the aftereffect is simply due to the difference in color of the faces, and not due to the difference in race in itself. In addition, the relationship between the adapting and test face was held constant within each block. This blocked presentation could have allowed participants to adopt different response strategies for the different types of trials. Experiment 2 was conducted to replicate the results of Experiment 1 while removing these two possible confounds. In Experiment 2 subjects viewed the same stimuli as in Experiment 1. However, the stimuli were presented in gray scale, so that a race-switch between the adapting and test face within

a trial did not provoke a difference in coloring between the two faces. In addition the different conditions (SI/SR, DI/SR, DI/DR) were mixed within blocks, so that participants could not build an expectation about the type of test face that would encounter.

METHODS

Participants

Twenty-four participants (average age of 21.8 years, 15 women) participated in Experiment 2. Of these 24 participants one was African-American, 17 were European-American, three were Hispanic, one was Middle Eastern, and two were Asian.

Stimuli

The faces used in Experiment 1 were altered to gray scale images using Adobe Photoshop. In all other regards, the stimuli in Experiment 1 and 2 were identical.

Procedure

The set-up was similar to Experiment 1, except that the three different adaptation conditions (SI/SR, DI/SR, and DI/DR) were presented randomly intermixed, divided in three blocks of 52 trials.

Analysis

The data were analyzed in the same way as in Experiment 1.

RESULTS

As before, subjects showed a strong emotion aftereffect [$F(1,23) = 193.1$, $p < 0.001$, partial $\eta^2 = 0.89$], which was again moderated by the relationship between the adapting and test face [$F(2,40) = 3.9$, $p = 0.04$, partial $\eta^2 = 0.14$]. The aftereffect was smaller in DI/DR trials compared to both the SI/SR [$t(23) = 2.1$, $p = 0.05$, Cohen's $d = 0.65$] and DI/SR trials [$t(23) = 2.2$, $p = 0.04$, Cohen's $d = 0.60$], whereas the aftereffect for SI/SR and DI/SR trials was comparable [$t(23) = 0.5$, $p = 0.60$, Cohen's $d = 0.15$], see **Figure 3**. Once more, these results show that a switch in race between the adapting and test face reduces the magnitude of the aftereffect, and this effect cannot be explained by a difference in face identity.
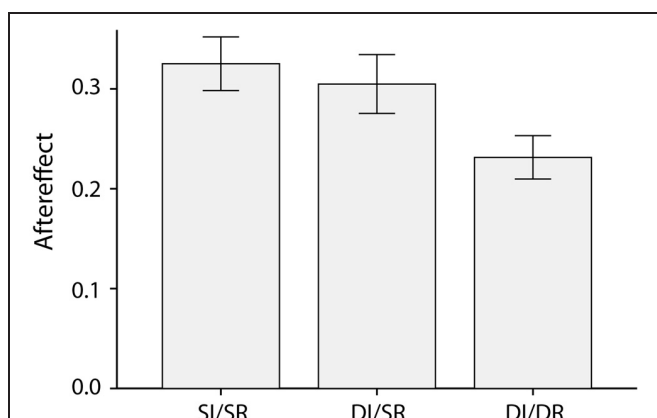

FIGURE 2 | Average emotion aftereffects, calculated as the difference between the proportion of ambiguous test-faces categorized as afraid following an angry adapting and a frightened adapting face, for the three different combinations of adapting and test face (Same Identity/Same Race, Different Identity/Same Race and Different Identity/Different Race).
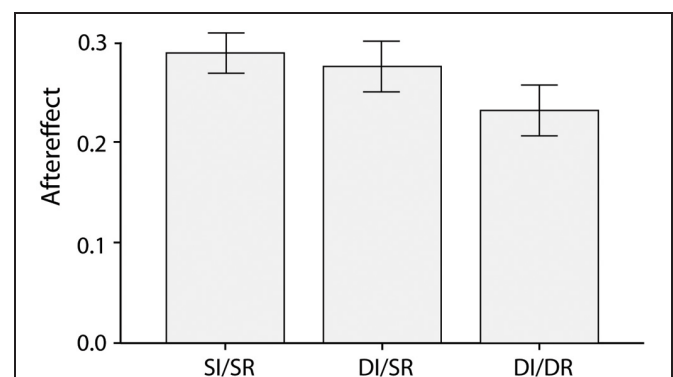

FIGURE 3 | Average emotion aftereffects for the three different combinations of adapting and test face (Same Identity/Same Race, Different Identity/Same Race, and Different Identity/Different Race) when the adapting and test faces were presented in gray scale.

DISCUSSION

These results provide an important indication that variable facial features such as emotional expression and invariable facial features such as social category are jointly processed. After adapting to a white face, subjects showed a smaller emotion aftereffect if the ambiguous test face was black (DR) than when the test face was white (SR). When the adapting and test face only differed in identity and not in race, aftereffects were comparable to trials in which the adapting face and the test faces had the same identity. This indicates that the neural representations of emotions expressed by white faces differ from the neural representations for the same emotions expressed by black faces. In addition, this difference in representation is not a consequence of differences in identity.

It thus seems that there is indeed a neural basis for the behavioral research showing that perception of variable cues like race is influenced by emotional expression and vice versa (Elfenbein and Ambady, 2003; Hugenberg and Bodenhausen, 2003, 2004; Ackerman et al., 2006; Richeson and Trawalter, 2008). Kubota and Ito (2007) have argued that the interdependence of race and emotion in these behavioral tasks only occurred because the stimuli were ambiguous in nature. In the current experiments, the test faces were indeed ambiguous in nature. However, it seems highly unlikely that the observed differences between the adaptation conditions are due to the ambiguity in the test faces. The observed effects hinge on the processing and adaptation of the *adapting* faces, and those faces were not ambiguous in race cues or emotion cues. In addition, participants were not required to respond to, or engage with the adapting faces in any way other than to just watch them for 5 s. It thus seems unlikely that participants employed specific strategies to deal with ambiguous faces when they processed the adaptive faces. In addition, many of the tasks that show behavioral interactions between emotion and race in face perception have used natural, unambiguous faces (Elfenbein and Ambady, 2003; Ackerman et al., 2006; Richeson and Trawalter, 2008). Taken together, the present and previous studies strongly suggest that emotion and social category information like race are jointly processed, and that the interactions at the behavioral level have their roots in the interdependent neural representation of race and emotional expressions.

These data furthermore suggest that the encoding of variable and invariable facial cues is more interconnected than has been assumed. Even though previous findings suggest that racial and emotional cues are processed independently (Bruce and Young, 1986; Haxby et al., 2000), at least in early neural processing stages

(Kubota and Ito, 2007), the current results indicate that neural representation of a face not only integrates variable and invariable cues such as race and emotion, but that the neural representations of facial expressions are significantly altered depending on the race of the face. This strongly suggests that race and emotion are processed interdependently. This corroborates a recent theoretical model of person perception by Freeman and Ambady (2011) which proposes that the perception of other people relies on an interplay of lower-level sensory perception and higher-order social cognition.

However, currently there are no studies (as far as we are aware) that provide evidence where in the brain the processing of emotion and race comes together. Previous research indicates that the neural mechanism that supports the integration of information about race and emotion does not appear to be cortical (Kubota and Ito, 2007). This suggests that subcortical structures, such as the amygdala, are a more likely site for the neural integration of race and emotion. It is well documented that the amygdala plays an important role in the processing of emotions, especially fear (Morris et al., 1996, 1998; Phelps, 2006). In addition, the amygdala is also implied in the processing of race cues in faces (Golby et al., 2001; Phelps et al., 2000). The fact that the pathways that underlie emotion and race processing cross at the amygdala provide suggestive evidence that the interaction between race- and emotion-related perceptual processes could take place here. Future brain-imaging studies will hopefully shed more light on the actual brain-structures that underlie the interaction between variable and invariable cues in face processing.

CONCLUSION

The race of the people around us has been previously shown to influence many socially relevant processes. The current findings indicate that race is indeed such a fundamental perceptual cue that it changes the basic perception of emotional facial expressions, which is in itself a highly relevant social cue. These results specifically show that racial cues exert their influence at a neural level, shaping the actual representation and perception of the emotion, and not just the subsequent interpretation of that emotion.

ACKNOWLEDGMENTS

REFERENCES

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., Maner, J. K., and Schaller, M. (2006). They all look the same to me (unless they're angry). *Psychol. Sci.* 17, 836.

Bestelmeyer, P. E. G., Jones, B. C., Debruine, L. M., Little, A. C., and Welling, L. L. M. (2010). Face aftereffects suggest interdependent

processing of expression and sex and of expression and race. *Vis. Cogn.* 18, 255–274.

Bijlstra, G., Holland, R. W., and Wigboldus, D. H. J. (2010). The social face of emotion recognition: evaluations versus stereotypes. *J. Exp. Soc. Psychol.* 46, 657–663.

Blau, V. C., Maurer, U., Tottenham, N., and McCandliss, B. D., (2007). The face-specific N170 component

is modulated by emotional facial expression. *Behav. Brain Funct.* 3, 7.

Bruce, V., and Young, A. (1986). Understanding face recognition. *Br. J. Psychol.* 77, 305–327.

Elfenbein, H. A., and Ambady, N. (2003). When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *J. Per. Soc. Psychol.* 85, 276–290.

Freeman, J. B., and Ambady, N. (2011). A dynamic interactive theory of

person construal. *Psychol. Rev.* 118, 247–279.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., and Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nat. Neurosci.* 4, 845–850.

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B. H., Marinkovic, K., and Rosen, B. R. (1999). Location of human face-selective cortex with

respect to retinotopic areas. *Hum. Brain Mapp.* 7, 29–37.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–232.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67.

Hoffman, E. A., and Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat. Neurosci.* 3, 80–84.

Hugenberg, K., and Bodenhausen, G. V. (2003). Facing prejudice: implicit prejudice and the perception of facial threat. *Psychol. Sci.* 14, 640–643.

Hugenberg, K., and Bodenhausen, G. V. (2004). Ambiguity in social categorization. *Psychol. Sci.* 15, 342.

Humphreys, G. W., Donnelly, N., and Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia* 31, 173–181.

Ito, T. A., and Bartholow, B. D. (2009). The neural correlates of race. *Trends Cogn. Sci.* 13, 524–531.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302.

Kubota, J. T., and Ito, T. A. (2007). Multiple cues in social perception: the time course of processing race and facial expression. *J. Exp. Soc. Psychol.* 43, 738–752.

Morris, J. S., Friston, K. J., Büchel, C., Frith, C. D., Young, A. W., Calder, A. J., and Dolan, R. J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain* 121, 47.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., and Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383, 812–815.

Mouchetant-Rostaing, Y., and Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. *J. Cogn. Neurosci.* 15, 900–910.

Phelps, E. A. (2006). Emotion and cognition: insights from studies of the human amygdala. *Psychology* 57.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* 12, 729–738.

Richeson, J. A., and Trawalter, S. (2008). The threat of appearing prejudiced and race-based attentional biases. *Psychol. Sci.* 19, 98.

Watson, T. L., and Clifford, C. W. G. (2003). Pulling faces: an investigation of the face-distortion aftereffect. *Perception* 32, 1109–1116.

Webster, M. A., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. *Nature* 428, 557–561.

Wu, J., Xu, H., Dayan, P., and Qian, N. (2009). The role of background statistics in face adaptation. *J. Neurosci.* 29, 12035.