



Intact performance on an indirect measure of race bias following amygdala damage

Elizabeth A. Phelps^{a,*}, Christopher J. Cannistraci^b, William A. Cunningham^c

^a Department of Psychology, New York University, 6 Washington Place, 8th Floor, New York, NY 10003, USA

^b Department of Radiology, Yale University Medical School, New Haven, CT, USA

^c Department of Psychology, Yale University, New Haven, CT, USA

Abstract

Recent brain imaging and lesion studies provide converging evidence for amygdala involvement in judgments of fear and trust based on facial expression [Adolphs et al., *Nature* 393 (1998) 470; Adolphs et al., *Neuropsychologia* 37 (1999) 1111; Breiter et al., *Neuron* 17 (1996) 875; Winston et al., *Nat. Neurosci.* 5 (3) (2002) 277]. Another type of social information apparent in face stimuli is social group membership. Imaging studies have reported amygdala activation to face stimuli of different racial groups [Hart et al., *NeuroReport* 11 (11) (2000) 2351]. In White American subjects, amygdala activation to Black versus White faces was correlated with indirect, implicit measures of racial evaluation [Phelps et al., *J. Cogn. Neurosci.* 12 (5) (2000) 729]. To determine if the amygdala plays a critical role in indirect social group evaluation, as suggested by the imaging results, a patient with bilateral amygdala damage and control subjects were given two measures of race bias. All subjects were female, White Americans. The Modern Racism Scale (MRS) is a direct, self-report measure of race attitudes and beliefs. The Implicit Association Test (IAT) is an indirect, automatic evaluation task. Performance on the two tasks did not differ between the patient with amygdala damage and control subjects. All subjects showed a pro-Black bias on the direct, explicit measure of race beliefs, the MRS, and a negative evaluation towards Black faces on the indirect measure of race evaluation, the IAT. These results indicate that even though amygdala activation to Black versus White faces is correlated with performance on indirect measures of race bias [Phelps et al., *J. Cogn. Neurosci.* 12 (5) (2000) 729], the amygdala is not critical for normal performance on the IAT. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Amygdala; Race; Evaluation; Automaticity

1. Introduction

The amygdala is a small, almond-shaped structure in the medial temporal lobe that is primarily known for its role in emotional learning and memory [18]. Across a wide range of species, the amygdala has been shown to be necessary for the acquisition and expression of aversive conditioning [10,17]. It has also been implicated in the modulation of memory with arousal [20]. However, more recently, investigations of the human amygdala have suggested that it may play a role in a limited range of social judgments as well, primarily judgments derived from facial stimuli.

In a series of studies, Adolphs et al. [2] have shown that damage to the amygdala leads to impairments of judgments of fear from facial expressions [29]. These studies have demonstrated that patients with amygdala damage, although they can often identify a fear expression, do not find this expression nearly as “fearful” as do normal controls. Func-

tional imaging studies have provided converging evidence for a specific role of the amygdala in the processing fear facial expressions. A number of studies have demonstrated greater amygdala activation to fear, relative to other facial expressions [7,21]. This differential amygdala response to fear expressions has also been observed when the faces are presented so quickly subjects are unaware of their presentation, demonstrating that the amygdala’s response is automatic and not dependent on conscious, control processes [27].

More recently both lesion and imaging studies have suggested amygdala involvement in more subtle social judgments. Investigations of patients with amygdala damage have found deficits in the ability to judge whether an individual appears to be approachable or trustworthy [1]. When shown pictures of individuals that normal control subjects rate as appearing untrustworthy or unapproachable, patients with amygdala damage tend to rate them as both trustworthy and approachable. Patients with amygdala lesions fail to pick up on the subtle differences in facial expression or appearance that normal subjects use to judge

* Corresponding author. Tel.: +1-212-998-8337; fax: +1-212-995-4249.
E-mail address: liz.phelps@nyu.edu (E.A. Phelps).

that an individual may not be trustworthy or friendly. Additional support for amygdala involvement in the complex judgment of trust from faces comes from an fMRI study by Winston et al. [28]. Subjects were asked to view faces and make a judgment of school age (high school or college) or trust (trustworthy or untrustworthy). After scanning, all of the faces were rated for trust. Winston et al. [28] reported greater amygdala activation to those faces rated as untrustworthy. This amygdala activation to untrustworthy faces was found regardless of the task performed while scanning (school or trust) suggesting that the amygdala response is automatic and not dependent on cognitive mediation. There was also a correlation between the trust ratings and facial expression, with faces rated as untrustworthy appearing more sad or angry. These results indicate that the amygdala is involved in the automatic judgment of personal qualities from facial expression.

The studies described above suggest a role for the amygdala in making social judgments about the characteristics of individuals from their facial *expression*; is the person fearful, trustworthy or approachable. Another variable used to judge personal characteristics from facial stimuli is *social group membership*. The determination that a person is male or female, young or old, Black or White may result in assumptions concerning that individual's personal qualities. Recent functional imaging studies examining the neural basis of social group evaluation have indicated amygdala involvement when viewing faces of different racial groups. Two fMRI studies have reported activation of the amygdala in response to the presentation of Black and White faces with neutral facial expressions [15,24]. In a study by Hart et al. [15], Black and White American subjects viewed pictures of Black and White faces. Overall, there was greater amygdala activation to outgroup faces than ingroup faces. In other words, Black American subjects showed more amygdala activation to White than Black faces, while the White American subjects showed the opposite pattern of results. This study demonstrated amygdala activation to social group membership, but did not indicate any potential behavioral role for this amygdala response.

A study by Phelps et al. [24] with White American subjects showed that the extent of amygdala activation to Black versus White faces was correlated with some indirect behavioral indications of racial group evaluation which were assessed after scanning. There is a growing body of behavioral evidence indicating a dissociation between some indirect (implicit) and direct (explicit) assessments of race evaluation. In spite of a decline over the past several decades of prejudicial attitudes toward Black and White social groups as measured by explicit self-report [6,25], there is robust evidence of racial bias using indirect assessments that bypass awareness and conscious control [5,11,14]. In the Phelps et al. [24] study the White American subjects who showed greater race bias on two indirect measures of race evaluation also showed greater amygdala activation while viewing the Black versus White faces. There was no

relation between amygdala activation and race bias as measured by explicit self-report. These results begin to delineate the different neural systems underlying direct and indirect evaluations of racial groups, and suggest that the amygdala may be involved in the automatic, implicit evaluation of social group information derived from faces.

The finding that amygdala activation is related to the indirect evaluation of social group information from facial stimuli is consistent with studies indicating indirect or automatic processing of other types of social information from faces [27,28]. However, the findings demonstrating a role for the amygdala in fear and trust perception are supported by converging evidence from brain imaging and lesion studies. The demonstration of activation of a particular brain region suggests that this region may be involved in the processing of the stimuli or behavior executed. But imaging data by itself cannot indicate the precise role that a brain region may have in a given task. Lesion studies can indicate if a specific brain region is critical for a given behavior, although like imaging studies there can be difficulties in interpreting lesion results [22]. Combining the two techniques can provide powerful converging evidence that activity in a brain region is not only correlated with stimulus processing, but critical for the execution of behavior.

In the present study we attempt to determine if the amygdala plays a critical role in the indirect evaluation of race bias. The correlation between amygdala activation and indirect measures of race bias suggests that the amygdala is engaging in the automatic processing of social group information from facial stimuli. However, this correlation does not indicate how the amygdala may be involved in the indirect evaluation of racial groups. By examining performance of a patient with bilateral amygdala damage on direct and indirect measures of race bias, we can determine if the amygdala is critical for normal performance on these tasks.

The measures of racial bias used were the same as those reported in the brain imaging study by Phelps et al. [24]. The Implicit Association Test (IAT) [14] was used to indirectly measure race bias. The term "bias" in this context refers to the presence of an indirect or non-controllable behavioral response that exhibits preference for one group over another. The IAT measures the degree to which social groups are automatically associated with positive and negative evaluations. The IAT was followed by the Modern Racism Scale, a commonly used measure of conscious, self-reported beliefs and attitudes toward Black Americans [19].

2. Methods

2.1. Subjects

A patient with bilateral damage to the amygdala (SP) was assessed along with two normal control subjects. SP is a 58-year-old White female who, at the age of 48 years,

had her right amygdala removed as a result of anteromedial temporal lobe resection for medically intractable epilepsy. Her right temporal lobe resection included partial removal of the middle and inferior temporal gyri, and complete removal of the hippocampus and parahippocampus. Prior to her surgery, an additional lesion was observed in the left amygdala. Two biopsies of this region revealed reactive gliosis consistent with mesial temporal sclerosis. Post-surgery T1 and T2 magnetic resonance scans show abnormal signal intensity throughout her left amygdala [23]. Neuropsychological and radiological indices suggest that the damage does not extend to adjacent temporal lobe structures in the left hemisphere. SP received a high school education, has taken college courses and presents a normal neuropsychological profile (see [23] for more details). Both control subjects were White females who were high school graduates and had taken some college courses. Their mean age was 65 years (67 and 63 years). Neither control subject had any significant medical history.

2.2. Stimuli

During the Implicit Association Test (IAT), subjects were presented photographs of nine Black and nine White male faces with neutral facial expressions. The photographs were taken from college yearbooks. All photographs were in black and white and depicted men with short hair, no facial hair and no distinctive clothing.

2.3. Procedures

Subjects were given two behavioral tests, one an indirect assessment of racial evaluation, and the other a direct, self-report measure of racial attitudes. The first was the Implicit Association Test (IAT). Subjects were asked to categorize Black or White male faces, while simultaneously categorizing words as good (*joy, love, peace*) or bad (*cancer, bomb, devil*). For half of the trials, subjects were asked to press a right button if the stimulus was either a White face or a good word and a left button if the stimulus was either a Black face or a bad word. For the remaining half of the trials, the pairings were reversed. The two conditions were counterbalanced. The difference in speed to respond to the Black + good/White + bad pairings compared to the Black + bad/White + good pairings provided the indirect measure of group evaluation. There were 60 trials of each pairing and the first 20 trials were considered practice and not included in further analysis. Outliers were defined as trials with reaction times were <300 ms or >3000 ms. Reaction times above this range were converted to 3000 ms. Trials with RT's below this range were eliminated. For all participants, fewer than 5% of trials were defined as outliers. Several studies using the IAT have now shown negative evaluation among White Americans in the form of faster responding in the Black + bad/White + good pairings [4,8,9,14] (for a demonstration of selected IAT procedures

visit www.yale.edu/implicit). SP performed this task twice in order to obtain a reliable measure of performance.

To assess racial bias directly by self-report, subjects completed the Modern Racism Scale (MRS). The MRS is a commonly used measure of conscious, self-reported beliefs and attitudes toward Black Americans [19]. Examples of items are: "Discrimination against Blacks is no longer a problem in the United States;" "It is easy to understand the anger of Black people in America." Scores on a six-point scale asking for agreement or disagreement with items were computed, with lower scores representing pro-Black and larger scores representing anti-Black beliefs and attitudes. This scale is a standard measure of attitudes and beliefs about the current status and rights of Black Americans and does not tap purely evaluative responses toward the group.

3. Results

3.1. Modern Racism Scale

SP's mean score for the items on the MRS was 1.8 and the mean for the two control subject was 1.2 (1.4 and 1). Both of these scores indicate pro-Black beliefs (with 6 being strongly anti-Black and 1 being strongly pro-Black), suggesting that SP and control subjects show a similar bias on this self-report measure of race attitudes.

3.2. Implicit Association Test

The measure of bias on the IAT is the discrepancy in reaction time between the Black + good/White + bad trials and the Black + bad/White + good trials. As can be seen in Fig. 1A, both SP and control subjects took longer to respond on the Black + good/White + bad trials. For SP the mean RTs (reported in ms) for Black + good/White + bad trials was 1756 (S.D. = 643) for Test 1 and 1860 (731) for Test 2. SP's mean RTs for Black + bad/White + good trials was 1132 (464) for Test 1 and 1099 (391) for Test 2. Control subjects also showed a discrepancy in mean RT to the two trial types. For the two control subjects the mean RTs for the Black + good/White + bad trials was 1071 (597) and 951 (296). The mean RT's for the Black + bad/White + good trials were 800 (382) and 677 (106), respectively. For all participants, the mean RTs for trials where a decision was made based on a word stimulus were slightly, but not significantly longer than for trials on which decisions were based on a face stimulus (mean for SP: words = 1617, faces = 1306; mean for controls: words = 890, faces = 856). These results are consistent with previous studies using White American subjects and suggest an indirect, negative anti-Black or pro-White evaluation [4,9].

However, relative to control subjects, SP took longer to respond overall and the discrepancy between her reaction times for the two trial types was greater than controls (693 ms difference for SP and 273 ms for controls).

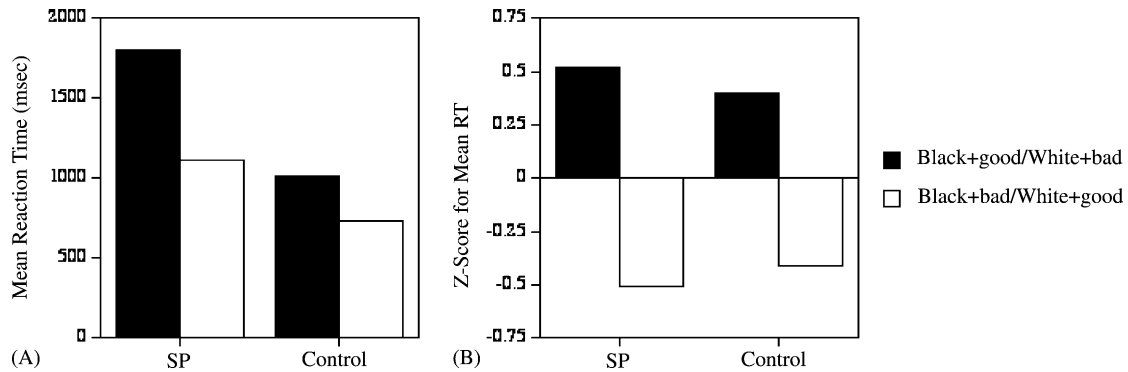


Fig. 1. (A) Mean reaction time to the Black + good/White + bad pairings and the Black + bad/White + good pairings for SP and normal control subjects; (B) z-scores of mean reaction times for SP and controls.

The overall response for control subjects was comparable to other studies using this task [4,8,9,24], however, SP's response times were outside the normal range. It not uncommon for brain injured patients to show slowed reaction times overall [23]. In an effort to compare SP's performance with control subjects using a standardized baseline, the mean and standard deviation of reaction times for all trials was calculated and used as a baseline to generate z-scores for the mean reaction times for the different trial types. The comparison of z-scores, presented in Fig. 1B, indicates that both SP and controls showed to a similar degree relatively faster reactions time for the Black + bad/White + good pairings ($z = -0.51$ for SP and -0.41 for controls) and a slower reaction times for the Black + good/White + bad pairings ($z = 0.52$ for SP and 0.40 for controls).

4. Discussion

The patient with bilateral amygdala damage performed similarly to control subjects on both the direct (Modern Racism Scale) and indirect (Implicit Association Test) measures of race evaluation. As in the Phelps et al. [24] imaging study, all the White American subjects in the present study showed a pro-Black bias on the direct, self-report measure of race attitudes and beliefs. They also showed a negative bias towards Black faces on the indirect assessment of race evaluation. Although it appears that SP may have shown an even greater negative bias on the IAT, based on the difference in reaction time to the Black + good/White + bad pairings and Black + bad/White + good pairings, when the reaction times were scaled to control for baseline differences, her bias response was similar to control subjects. This discrepancy between pro-Black beliefs and attitudes as measured by direct, self-report and an anti-Black, pro-White bias as assessed with the IAT is consistent with other studies with White American subjects [9,14,24].

The finding that SP demonstrates a negative bias towards Black faces using an indirect assessment of race evaluation

suggests that the amygdala is not critical for the indirect expression of race bias. Unlike the studies showing that amygdala damage leads to impairments in judgments of fear or trust from facial expression [1,2], the amygdala does not appear to be necessary for the indirect evaluation of social group membership, at least as measured by the IAT.

The question remains as to what role the amygdala may play, if any, in social group evaluation. The present study, when viewed in light of the correlation observed between IAT performance and amygdala activation in the Phelps et al. [24] study, suggests two preliminary conclusions. First, the amygdala response to social group information derived from faces appears to be automatic. In the Phelps et al. [24] study, the task performed during scanning did not require any judgments related to race and the subjects were unaware that race was a variable being investigated until after scanning. Second, this automatic response of the amygdala to racial group membership and race bias as measured by the IAT are both the result of similar evaluative mechanisms whose neural substrates are, at this time, unknown.

Given what is known about the neural systems of social group face processing, one possibility is that the amygdala activation response to Black versus White faces is related to differences in the perceptual processing of same versus other race faces. It has been demonstrated that for both White and Black Americans, outgroup, or other race faces will lead to less activation of the fusiform face area (FFA) [13]. The FFA is a region of the temporal lobe known to be involved in the perception of faces [16]. High level visual cortices in the temporal lobe project to the amygdala [3]. It has been suggested that the FFA responds to greater expertise in processing faces [12], and that we tend to be less "expert" at processing other race faces. In other words, the diminished FFA response to other race faces may be the result of less expertise in processing outgroup faces [13]. It is possible that the impoverished neural processing of other race faces leads to some subtle ambiguity or uncertainty in response to these faces. The amygdala may be particularly sensitive to uncertainty in the environment because it may signal potential

danger [26]. This would suggest that as we become more “expert” at processing other race faces we might expect more FFA activation and less amygdala activation. Consistent with this hypothesis, Phelps et al. [24] did not observe amygdala activation to highly familiar Black versus White faces.

Any link between IAT performance and expertise or the FFA response is less clear. The IAT is a task that has been used with a range of stimuli, not just faces, suggesting that performance on this task is mediated by factors other than face processing [8,9,14]. In addition, familiarity with stimuli does not eliminate biased performance on the IAT [4,24]. Nevertheless, expertise requires more than familiarity and it is possible that expertise with other race exemplars would help moderate the evaluative judgment that underlies the IAT.

It is important to note that the lack of a deficit on performance of the IAT following amygdala damage does not rule out any critical role for the amygdala in the indirect evaluation of racial groups. The IAT is a categorization task. The bias in reaction time that results from indirect evaluation in the IAT may not be amenable to conscious control, but performing the task itself requires cognitive mediation. Indirect evaluation tasks of facial stimuli that are less dependent on cognitive mediation could be impaired following amygdala lesions. In addition, the fact that the amygdala is not necessary for normal levels of performance on the IAT does not rule out a role for the amygdala in normal performance. It is possible that SP is processing the stimuli differently than normal controls to compensate for her lesion when performing this task. Finally, these results do not rule out role for the amygdala in acquisition of race bias responses. Even though the amygdala is not necessary to express an indirect race bias as assessed by the IAT, it may contribute to the acquisition such responses. Data from additional patients with amygdala damage and/or additional race bias tasks would be helpful in addressing some of these issues.

As yet, it is not possible to determine the common neural pathways related to both the indirect evaluation of race groups, as measured with the IAT, and amygdala activation to faces of different races. It is clear that the amygdala can respond to social group information portrayed in faces [15,24]. However, it is not clear what quality of the face stimuli drives this amygdala response or what behavioral role, if any, this differential amygdala response may have.

Acknowledgements

The studies were conducted at the Yale University School of Medicine. The authors would like to thank Benjamin Holmes and Laura Thomas for assistance in the preparation of this manuscript. This research was supported by the James S. McDonnell Foundation, 21st Century Scientist Award to EAP.

References

- [1] Adolphs R, Tranel D, Damasio AR. The human amygdala in social judgement. *Nature* 1998;393:470–4.
- [2] Adolphs R, Tranel D, Hamann S, Young AW, Calder AJ, Phelps EA, et al. Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia* 1999;37:1111–7.
- [3] Amaral DG. Anatomical organization of the primate amygdaloid complex. In: Aggleton JP, editor. *The amygdala: neurobiological aspects of emotion, memory, and mental dysfunction*. New York, NY: Wiley; 1992. p. 1–66.
- [4] Banaji MR. Implicit attitudes can be measured. In: Roediger III HL, Narine JS, editors. *The nature of remembering: essays in honor of Robert G. Crowder*. Science conference series. Washington, DC: American Psychological Association; 2001. p. 117–50.
- [5] Bargh JA, Chen M. Non-conscious behavioral confirmation processes: the self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology* 1997;33: 541–60.
- [6] Biernat M, Crandall CS. Racial attitudes. In: Robinson PH, Shaver D, Wrightsman LS, editors. *Measures of political attitudes*. San Diego, CA: Academic Press; 1999. p. 291–412.
- [7] Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, et al. Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 1996;17:875–87.
- [8] Cunningham WA, Preacher KJ, Banaji MR. Implicit attitude measure: consistence, stability and convergent validity. *Psychological Science* 2001;12(2):163–70.
- [9] Dasgupta N, McGhee DE, Greenwald AG, Banaji MR. Automatic preference for White Americans: eliminating the familiarity explanation. *Journal of Experimental Social Psychology* 2000;36(3): 316–28.
- [10] Davis M. Neurobiology of fear responses: the role of the amygdala. *Journal of Neuropsychiatry and Clinical Neurology* 1997;9:382–402.
- [11] Fazio RH, Jackson JR, Dunton BC, Williams CJ. Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology* 1995;69:1013–27.
- [12] Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. Activation of the middle fusiform face area increases with expertise in recognizing novel objects. *Nature Neuroscience* 1999;2:568–73.
- [13] Golby AJ, Gabrieli JDE, Chiao JY, Eberhardt JL. Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience* 2001;4(8):845–50.
- [14] Greenwald AG, McGhee JL, Schwartz JL. Measuring individual differences in social cognition: the Implicit Association Test. *Journal of Personality and Social Psychology* 1998;74:1464–80.
- [15] Hart AJ, Whalen PJ, Shin LM, McInerney SC, Fischer H, Rauch SL. Differential response in the human amygdala to racial outgroup versus ingroup face stimuli. *NeuroReport* 2000;11(11):2351–5.
- [16] Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 1997;17:4302–11.
- [17] Kapp BS, Whalen PJ, Supple WF, Pascoe JP. Amygdaloid contributions to conditioned arousal and sensory information processing. In: Aggleton JP, editor. *The amygdala: neurobiological aspects of emotion, memory, and mental dysfunction*. New York, NY: Wiley; 1992. p. 471–84.
- [18] LeDoux JE. *The emotional brain*. New York, NY: Simon & Schuster; 1996.
- [19] McConahay JP. Modern racism, ambivalence, and the Modern Racism Scale. In: Dovidio JF, Gaertner SL, editors. *Prejudice, discrimination, and racism*. Orlando, FL: Academic Press; 1986. p. 91–125.
- [20] McGaugh J. Memory: a century of consolidation. *Science* 2000;287(5451):248–51.
- [21] Morris JS, Frith CD, Perrett DI, Rowland D. A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 1996;383:812–5.

- [22] Phelps EA. Brain versus behavioral studies in cognition. In: Sternberg R, editor. *Concepts in cognition*. Cambridge, MA: MIT Press; 1999.
- [23] Phelps EA, LaBar KS, Anderson AK, O'Conner KJ, Fulbright RK, Spencer DD. Specifying the contributions of the human amygdala to emotional memory: a case study. *Neurocase* 1998;4:527–40.
- [24] Phelps EA, O'Conner KJ, Cunningham WA, Funayama ES, Gatenby JC, Gore JC, et al. Performance in indirect measure of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience* 2000;12(5):729–38.
- [25] Schuman H, Steeh C, BoBo L. *Racial attitudes in America: trends and interpretations*. Cambridge, UK: Harvard University Press; 1997.
- [26] Whalen PJ. Fear, vigilance, and ambiguity: initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science* 1998;7:177–88.
- [27] Whalen PJ, Rauch SL, Etcoff NL, McInerney SC, Lee MB, Jenike MA. Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience* 1998;18:411–8.
- [28] Winston JS, Strange BA, O'Doherty J, Dolan RJ. Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience* 2002;5(3):277–83.
- [29] Young AW, Aggleton JP, Hellawell DJ, Johnson M. Face processing impairments after amygdalotomy. *Brain* 1995;118:15–24.