

This article was downloaded by: [Vrtika, Pascal]

On: 27 July 2009

Access details: Access Details: [subscription number 913435714]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Social Neuroscience

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t741771143>

Memory for friends or foes: The social context of past encounters with faces modulates their subsequent neural traces in the brain

Pascal Vrtika ^a; Frédéric Andersson ^b; David Sander ^c; Patrik Vuilleumier ^a

^a University of Geneva, and University Medical Center of Geneva, Geneva, Switzerland ^b University Medical Center of Geneva, Geneva, Switzerland ^c University of Geneva, Geneva, Switzerland

First Published on: 27 July 2009

To cite this Article Vrtika, Pascal, Andersson, Frédéric, Sander, David and Vuilleumier, Patrik(2009)'Memory for friends or foes: The social context of past encounters with faces modulates their subsequent neural traces in the brain', Social Neuroscience,99999:1,

To link to this Article: DOI: 10.1080/17470910902941793

URL: <http://dx.doi.org/10.1080/17470910902941793>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Memory for friends or foes: The social context of past encounters with faces modulates their subsequent neural traces in the brain

Pascal Vrtička

University of Geneva, and University Medical Center of Geneva, Geneva, Switzerland

Frédéric Andersson

University Medical Center of Geneva, Geneva, Switzerland

David Sander

University of Geneva, Geneva, Switzerland

Patrik Vuilleumier

University of Geneva, and University Medical Center of Geneva, Geneva, Switzerland

Every day we encounter new people, interact with them, and form person impressions based on quick and automatic inferences from minimal contextual information. Previous studies have identified an extensive network of brain areas involved in familiar face recognition, but there is little evidence to date concerning the neural bases of negative vs. positive person impressions. In the present study, participants were repeatedly exposed to 16 unfamiliar face identities within a pseudo-interactive game context to generate a perception of either “friends” or “foes”. Functional magnetic resonance imaging (fMRI) was then performed during an old/new memory task to assess any difference in brain responses to these now familiar face identities, relative to unfamiliar faces. Importantly, whereas facial expressions were always emotional (either smiling or angry) during the encoding phase, they were always neutral during the memory task. Our results reveal that several brain regions involved in familiar face recognition, including fusiform cortex, posterior cingulate gyrus, and amygdala, plus additional areas involved in motivational control such as caudate and anterior cingulate cortex, were differentially modulated as a function of a previous encounter, and generally more activated when faces were perceived as “foes” rather than “friends”. These findings underscore that a key dimension of social judgments, based on past impressions of who may be supportive or hostile, may lead to long-lasting effects on memory for faces and thus influence affective reactions to people during a subsequent encounter even in a different (neutral) context.

Keywords: Familiar face recognition; Friends or foes; Social judgement; Affective reactions; fMRI.

Correspondence should be addressed to: Pascal Vrtička, Swiss Center for Affective Sciences, 7, rue des Batoirs, 1205 Geneva, Switzerland. E-mail: pascal.vrticka@unige.ch

This research was supported by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (no. 51NF40-104897) and hosted by the University of Geneva; and grants of the Swiss National Science Foundation to DS and PV.

INTRODUCTION

Our everyday life is marked by innumerable social interactions with a wide range of different people, including family members, friends, professional colleagues, as well as complete strangers. To determine how to interact appropriately with these various individuals, the ability to form and remember, even implicitly, general impressions about a given person is a crucial social skill of human beings.

Several studies have shown that only minimal information is needed to form person impressions from just seeing faces (Todorov & Uleman, 2002, 2003, 2004), or to make trait judgments such as trustworthiness, competence, and aggressiveness (Willis & Todorov, 2006). This social evaluation process may occur not only unintentionally but also very rapidly, since very short exposure times to faces are sufficient for some of these judgments (Bar, Neta, & Linz, 2006; Willis & Todorov, 2006). Besides trait inferences from seeing faces, other studies have also demonstrated that people can make quick unreflective inferences concerning a person based on minimal behavioral information, even when the behavior can be explained by situational context rather than personal dispositions (Gilbert & Malone, 1995; Trope & Alfieri, 1997). Moreover, it has been shown that people make such rapid inferences about another person even when they do not intend to do so (Uleman, Blader, & Todorov, 2005), and when their cognitive resources are constrained by a concomitant task (Bargh, 1994), suggesting that this impression formation process is fairly automatic.

Importantly, once person impressions are formed during social interactions, storing this information in memory is crucial to make use of these inferences and other person-related knowledge when a person is encountered again (see Mealey, Daood, & Krage, 1996; Oda, 1997; Takahashi, 2005). In particular, it has been shown that memory for person identities can be modulated at subsequent retrieval when these persons were first seen with some emotion expressed in the face (D'Argembeau & Van der Linden, 2007) or the voice (Armony, Chochol, Fecteau, & Belin, 2007), as compared with when their expression is neutral.

The neural mechanisms underlying these social processes are still poorly known, but there is increasing evidence that several brain regions are critically important for social appraisals. Among

these, the amygdala has been shown to be involved in tracking the trustworthiness of faces by responding to some categories of faces that have properties commonly perceived as signalling untrustworthiness (Adolphs, Tranel, & Damasio, 1998; Armony et al., 2007; Engell, Haxby, & Todorov, 2007; Winston, Strange, O'Doherty, & Dolan, 2002). Other regions involved in face processing such as the fusiform gyrus (FG) and superior temporal sulcus (STS) have been found to be modulated by socially significant signals, even during tasks that do not require any explicit social or affective inferences (Frith & Frith, 2003; Singer, Kiebel, Winston, Dolan, & Frith, 2004; Vuilleumier, Armony, Driver, & Dolan, 2001; Winston et al., 2002). In addition, prefrontal regions including the anterior cingulate cortex (ACC) may also be differentially activated as a function of perceived social rejection during exposure to faces (Somerville, Heatherton, & Kelley, 2006) or pseudo-interactive social games (Eisenberger, Lieberman, & Williams, 2003).

However, it still remains largely unknown how information from a first encounter with a person determines the outcome of future interactions with the same person, and how memory traces from past impressions may affect brain responses during subsequent encounters in a different context. Only a few recent studies have begun to investigate how person impressions based on facial and/or behavioral information may translate into more enduring memory representations, and later influence social processing during new encounters with familiar persons or faces.

A first approach widely used in early studies investigating the effects of social knowledge and memory of face perception was to compare brain responses to personally familiar or intimate persons, relative to famous and unknown faces (Bartels & Zeki, 2004; Shah et al., 2001). Results from these studies have provided the basis for an influential model of face recognition and social cognition (Gobbini & Haxby, 2007). This model proposed that recognition of familiar faces involves not only brain areas coding for invariant visual features of face identity (such as FG), but also brain regions associated with person knowledge (including STS, posterior cingulate cortex (PCC), and anterior temporal areas) as well as emotional processing (such as insula and amygdala). This model therefore implies that, besides the retrieval of purely visual traces stored from previous encounters, successful recognition of familiar faces involves the reactivation of several

other essential components including information about the personality, attitude, and outlook of a person, as well as the appropriate emotional response associated with this individual (Gobbini & Haxby, 2007).

Even though the model developed by Gobbini and Haxby (2007) provides a solid framework for the role of specific brain regions in these different aspects of familiar face recognition, it does not delineate the neural components involved in the formation and maintenance of different types of person impressions, such as basic distinctions between positive or negative impressions. Importantly, because the model was initially based on two studies comparing personally familiar faces with famous or unknown faces (Gobbini, Leibenluft, Santiago, & Haxby, 2004) or pictures of one's own children with familiar unrelated and unfamiliar children (Leibenluft, Gobbini, Harrison, & Haxby, 2004), neural activations imputed to knowledge and affect associated with familiar faces might have been biased (or actually limited) to positive traits. Another imaging study that manipulated the acquired moral status of previously unfamiliar faces, by using a pseudo-interactive economic game (Singer et al., 2004), also reported only activations evoked by pictures of cooperators versus defectors (hence related to inherently positive and rewarding values), but did not report the inverse comparison, concerning any response associated with negative impressions. Only a single recent study tested for the neural correlates of negative person traits by comparing activations to previously unknown faces after they were paired with scenarios describing neutral, nice, aggressive, and disgusting behaviors (Todorov, Gobbini, Evans, & Haxby, 2007). This study reported that the left STS was more activated for negative versus positive traits, which was attributed to a stronger engagement of processes underlying the analysis of social intentions (Frith & Frith, 1999) in the case of negative affective knowledge about a person. In addition, bilateral anterior insula was also found to show stronger responses to faces associated with aggressive and disgusting behaviors, consistent with the role of this region in processing disgust-related stimuli (Calder, Keane, Manes, Antoun, & Young, 2000).

In the present fMRI study, we therefore designed a new paradigm that allowed us to investigate the neural correlates of processing both positive and negative person traits in familiar face recognition, as formed implicitly following encounters in different social contexts. For

this purpose, we engaged our participants in a pseudo-social interaction game task with a set of unknown faces that were presented as virtual partners. Distinct social attitudes were generated for these faces by systematically manipulating the social significance of their facial expression in relation to different task contexts, so as to elicit a perception of either social alliance (i.e. friendly partners) or social hostility (i.e. opponent partners) during this pseudo-game phase (see Methods). Critically, the same face identities were later shown again during fMRI, now with a neutral expression but familiar to the participants and associated with different affective values. Importantly, because all face identities were presented with a neutral facial expression and without any social context during this second phase, any differences in brain responses to these faces should reflect acquired person knowledge based on the past encounters and previously formed person impressions.

We predicted that brain activations related to the familiarity of previously seen faces should show a distinct profile as a function of their social-emotional value, including for regions involved in face-processing, such as FG or STS, but also for regions involved in affective processing such as the amygdala, ACC, or orbitofrontal areas, which have previously been found to track (un)trustworthiness of faces (Winston et al., 2002), signs of social rejection (Somerville, Heatherton, & Kelley, 2006), or social attachment (Delgado, Frank, & Phelps, 2005; Singer et al., 2004), respectively.

METHODS

Subjects

We recruited 16 healthy paid volunteers (8 males, mean age 23.6 ± 3.6 , all right-handed) from the student population of Geneva University. All had a normal or corrected to normal vision, no history of neurological or psychiatric disease, and gave informed consent according to the local ethical committee regulation.

Stimuli and procedure

The experiment consisted of two phases. In the first phase, a series of faces were presented with different emotional expressions in different task

contexts to elicit distinct social appraisals associated with each face identity (encoding phase). Next, in the critical test phase, the same faces were presented again, together with new faces, for an explicit recognition task (memory phase).

Encoding phase

Participants were first exposed to a pseudo-interactive game context in which they were presented with faces from virtual partners as feedback on their performance. The game was a

visual dot-counting task, presented to participants as a test of perceptual speed.

Each trial began with a white central fixation-cross on a black screen (for 3 to 7 s, average 3.5 s), followed by a brief visual display divided in two parts with a variable number of white dots on each side of the screen (presented for 500 ms; see Figure 1a). The number of dots on each side ranged from 10 to 15. Their quantity and position were randomly assigned on every trial for each side separately, in such a way that the display was never visually identical on the two sides. Partici-

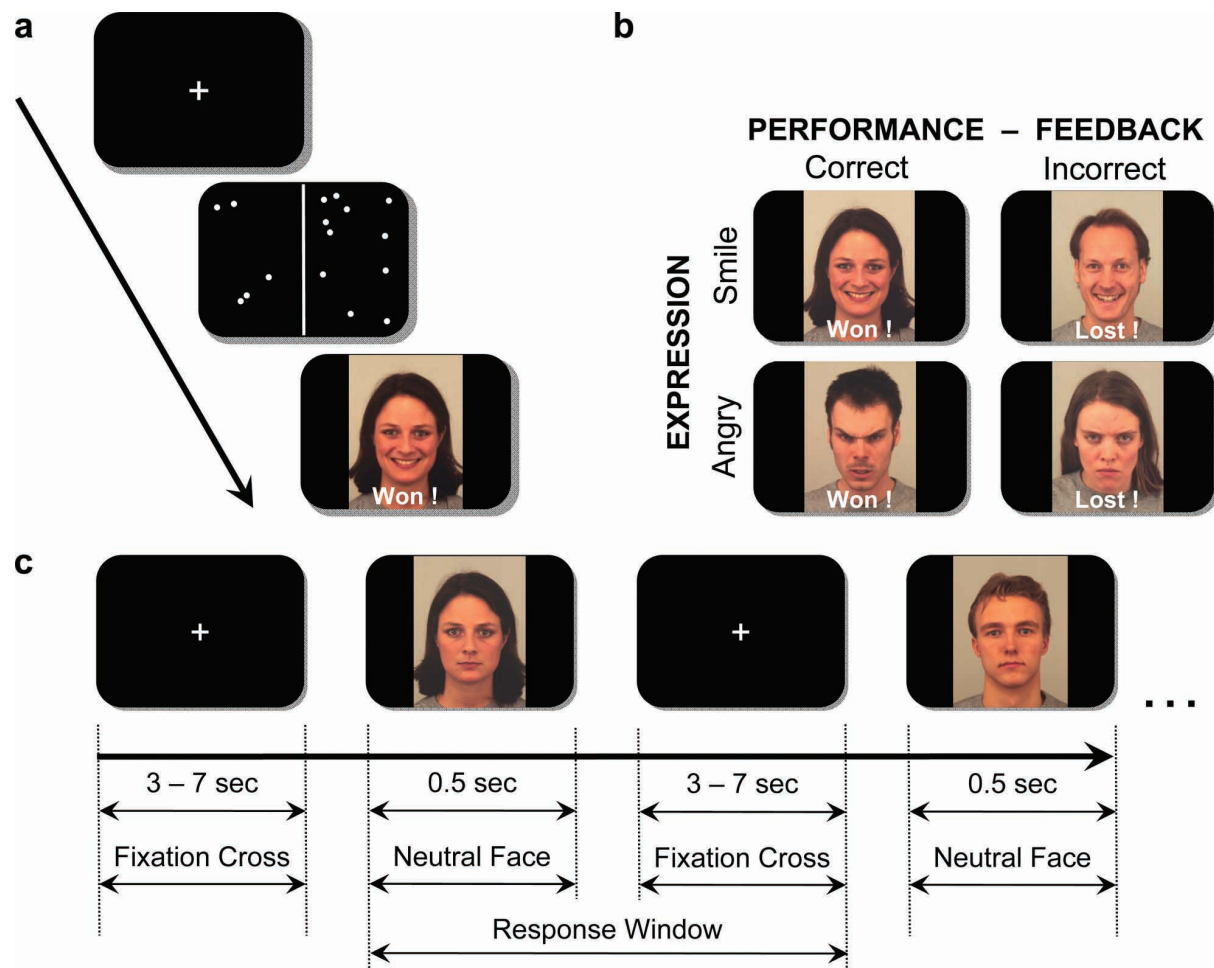


Figure 1. Illustration of the paradigm and the four different feedback conditions. (a) Experimental paradigm during the initial encoding phase. On each trial, participants first saw a central fixation cross, followed by the dot-counting task (0.5 s), in which they had to indicate the side of the screen that contained more dots (right vs. left). Following each response, a visual feedback was shown (1.5 s), composed of a word together with a face. (b) Illustration of the four different feedback conditions during encoding. Two socially “congruent” (Smiling Face on WON trial, Angry Face on LOST trial) and two socially “incongruent” (Smiling Face on LOST trial, Angry Face on WON trial) combinations were possible. Four different face identities (two female and two male) were used in each of these four conditions. (c) Experimental paradigm during the recognition phase. On each trial, participants first saw a central fixation cross, followed by a face with a neutral expression. They had to indicate whether each face was familiar (“old”) or unfamiliar (“new”). See Methods for further details.

pants had to indicate which side of the screen contained more dots (right/left) by pressing one of two response-keys. The total number of dots and the difference between the two display sides were adjusted online based on the participant's performance on the preceding trial, by reducing the difference after each correct trial (minimum 1 dot) or increasing the difference after each incorrect trial (maximum 5 dots), allowing us to maintain performance close to threshold and to obtain approximately equal numbers of correct and incorrect trials (mean correct = $57 \pm 1\%$ across conditions). In addition, to further ensure this equal distribution (crucial for a subsequent unbiased comparison of conditions), we inserted occasional "trick" displays with 15 dots on both sides whenever performance exceeded 60% correct of two consecutive trials (20 ± 5.6 out of 128 trials). None of the participants noticed these "impossible" trials. The dot display was then replaced by a black screen with a variable interval (jitter of 1000 to 1400 ms, average 1200 ms), during which participants gave their response.

Critically, each response was followed by a feedback screen (1500 ms), consisting of a face that could have either a smiling or an angry expression, together with a verbal indication of the actual outcome on the counting task (either "WON" or "LOST"; see Figure 1a). The latter verbal feedback always corresponded to the real performance (correct or incorrect) on the preceding trial (except on the few "trick" trials with equal number of dots on both sides, where a negative "LOST" feedback was given to reduce any excess of correct over incorrect trials). By contrast, the emotional expression of faces was pseudo-randomly assigned on every trial, with the constraints that smiling and angry faces were presented on an equal number of correct and incorrect trials each; and that a given face identity was always seen with the same expression (smiling or angry) and the same feedback message (WON or LOST) throughout the task for any given participant (with different face identities in the different conditions counterbalanced across participants). This design resulted in four different combinations of verbal and facial feedback (see Figure 1b), of which two were *congruent*: smiling face on WON trial (SF-W) or angry face on LOST trial (AF-L); and two were *incongruent*: angry face on WON trial (AF-W) or smiling face on LOST trial (SF-L).

We induced a pseudo-social interactive context by telling participants a cover-story along the following lines: faces were those of other subjects included in the study and belonged to two different teams; the participant was randomly assigned to play for the benefit of one team (allied), and hence against the other team (opponent); the goal of the study was to compare perceptual abilities and cooperation among the groups, and performance of the participant determined gains for the allied team but losses for the other opponent team when his/her responses were correct (and vice versa, incorrect performance of the participant led to losses for the allied team but gains for the other team). Hence, each correct response gave one point to the allied group, while each incorrect response gave one point to the other group. These outcomes would be reminded during the game by presenting faces from individuals in the different groups with appropriate expressions (either smiling or angry) together with feedback about the participant's performance (won or lost). Subsequent debriefing after scanning indicated that participants accepted the cover-story, were highly motivated by the task, and experienced an expected affective response including "irritation", or "satisfaction" as a function of the incongruent or congruent social meaning of the facial expression paired with feedback (see Vrtička, Andersson, Grandjean, Sander, & Vuilleumier, 2008). Importantly, different faces were presented for the different groups of allies and opponents, counterbalanced across participants.

All face stimuli were colour photographs taken from the Karolinska Directed Emotional Faces set (KDEF, Lundquist, Flykt, and Öhmann, 1998). We selected 16 different individuals (8 males) to be shown in the first phase (learning), while 16 others were shown only the second phase (memory recognition task, see below). Each face identity was assigned to one condition only (two male and two female in each of the four feedback types for faces seen in phase 1; plus eight male and eight female for the new unfamiliar faces subsequently seen in phase 2), with all identities counterbalanced across participants. During the pseudo-game task (phase 1), each of the 16 faces was shown 8 times in the corresponding condition, in random order, resulting in 128 trials in total per participant (total duration ~15 min).

Imaging results from this phase have been reported separately in relation to the influence

of individual personality traits on social behavior and brain response (Vrtička et al., 2008).

Memory phase

After a short break of ~5 min without any intervening task, participants were given a surprise recognition memory test, in which they had to distinguish between previously seen faces and new faces (“yes” or “no”). Each trial began with a white central fixation-cross on a black screen (for 3 to 7 s, average 3.5 s), followed by a brief visual display showing a single face with a neutral expression (presented for 500 ms). After its offset, participants had to press one of two response-keys to indicate whether the current facial identity had already been presented during the previous game phase (“old”: familiar face condition) or whether it was a novel face shown only during this second memory test phase (“new”: unfamiliar face condition; Figure 1c).

The familiar faces were from the 16 identities already displayed during the encoding phase (phase 1), with the important difference that now all faces were seen with a neutral expression (contrary to the first phase, where they were shown with angry or smiling expressions). Thus, although these faces were initially seen in four distinct conditions during the previous pseudo-game, all now were visually similar and required the same response (i.e. “old” judgement), so that only their previous “social history” could make them different from each other or from novel faces. The unfamiliar faces were 16 new identities (also with a neutral facial expression).

All faces (familiar and unfamiliar) were intermingled in a pseudo-randomized order, with the only constraint that there were never more than two successive presentations of new or old faces. Each face identity was repeated six times (in different successive order) during this phase, resulting in 192 trials in total per participant (total duration ~16 min). Participants were informed about these repetitions and instructed to respond “old” only to faces seen during the previous game phase. To avoid fatigue, the memory task was split in two scanning runs (each 8 min).

Eye-tracking

Eye movements were monitored continuously during both phases with an MRI-compatible infrared eyetracker LRO L6 (Applied Science Laboratories, Bedford, MA, USA). Eye position

coordinates (x and y) were recorded at 60 Hz and saved for offline analysis, allowing us to test for any difference in visual inspection of faces between different conditions, during either encoding or recognition.

Post-scanning memory assessment

After scanning, all participants were presented with a “likeability” and “sorting” rating task for “old” faces. For likeability ratings, subjects were shown pictures of these 16 faces (with a neutral expression) and asked to report how likeable each face appeared for them (on a five-point scale from -2 to 2). For the sorting task, subjects were again shown these 16 “old” faces (with a neutral expression), and now told explicitly about the four feedback conditions of the encoding phase; they were then asked to classify each face in the corresponding feedback condition in which it was initially presented (four faces in each of the four categories, forced choice). This allowed us to assess any explicit knowledge for the social context previously associated with each face.

MRI acquisition and analysis

MRI data were acquired on a 1.5 T whole-body INTERA system (Philips Medical Systems), using standard head-coil configuration. For each participant, a structural image was obtained with a 3D-GRE T1-weighted sequence (FOV = 250 mm, TR/TE/Flip = 15 ms/5.0 ms/30°, matrix = 256 × 256, slice-thickness = 1.25 mm) and functional images with a GRE EPI sequence (TR/TE/Flip = 2500 ms/40 ms/80°, FOV = 250 mm, matrix = 128 × 128). Functional images covered the whole brain, composed of 30 contiguous 4 mm axial slices parallel to the inferior edge of the occipital and temporal lobes, and acquired continuously for a total of 232 images per participant while they performed the recognition memory task (two successive runs of 116 images each).

Functional data were analyzed using the general linear model for event-related designs in SPM2 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm). All images were realigned, corrected for slice timing, normalized to an EPI template (resampled voxel-size of 3 mm), spatially smoothed (8 mm FWHM Gaussian kernel). A high-pass frequency filter (cutoff 120 s) and corrections for autocorre-

lation between scans were applied to the time series.

Statistical analysis was performed using the general linear model implemented in SPM2, with a separate regressor for each event type convolved with a canonical hemodynamic response function. Ten events from the recognition memory task were modeled, including correct and incorrect trials for the old faces previously seen in each of the four possible feedback conditions (SF-W, SF-L, AF-W, AF-L, either correctly recognized or missed); plus the correct and incorrect trials for the new faces (correct rejections and false alarms, respectively). Due to our splitting of data between correct and incorrect recognition, this distribution resulted in an average of ~ 15 events for the correct recognition of old faces in each condition. Movement parameters from realignment corrections were entered as additional covariates of no interest to account for residual movement artifacts after realignment. Statistical parametric maps were generated from linear contrasts between the different conditions in each participant. A second-stage random-effect analysis was then performed using one-sample t -tests on contrast images obtained in each subject for each comparison of interest. All contrasts were performed across the whole brain using standard threshold criteria (Worsley et al., 1996) of significant activation at a voxel-level of $p < .001$ uncorrected and cluster size greater than 5 voxels (135 mm^3). Average parameter estimates of activity (betas) for each feedback condition were extracted from all voxels in regions of interest (ROIs), defined by the full-extent clusters showing significant activation at a voxel-level of $p < .001$ (uncorrected) in the SPM group analysis (random-effect contrasts).

RESULTS

Behavioral data

Recognition memory during fMRI scanning required subjects to make “old” vs. “new” judgments on neutral faces corresponding to either familiar identities that were seen in different social context during the encoding phase, or unfamiliar identities that were seen only during the memory phase. Overall memory performance was well above chance (total mean correct: $71.54 \pm 1.65\%$) for both familiar and unfamiliar faces (Figure 2a). Correct rejection rates for unfamiliar faces were higher than correct recognition rates for familiar

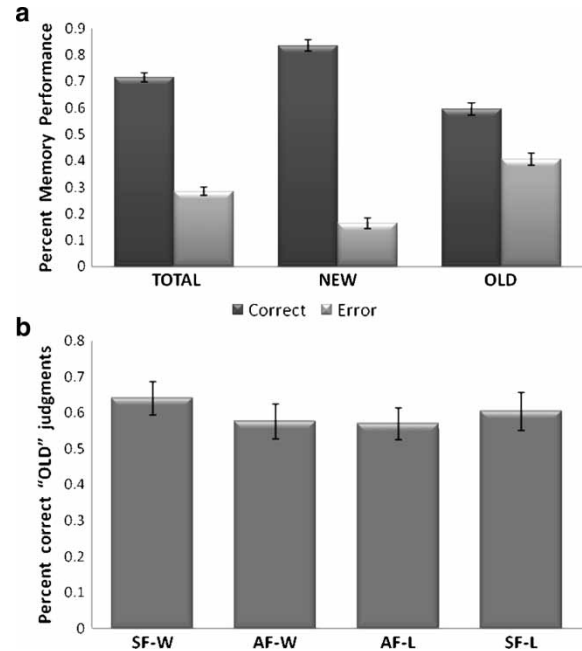


Figure 2. Behavioral memory performance during scanning. (a) Memory performance during the face recognition phase. TOTAL correct: $71.54 \pm 1.65\%$; NEW correct: $83.58 \pm 2.15\%$; OLD correct: $59.50 \pm 2.34\%$. (b) Memory performance across the different conditions of previous feedback context (SF-W: smiling face won, AF-W: angry face won, AF-L: angry face lost, SF-L: smiling face lost) showed no significant differences between the four face categories.

faces, $83.6 \pm 2.1\%$ and $59.5 \pm 2.3\%$ respectively; $t(15) = 7.852$, $p < .001$, paired t -test. However, there was no significant difference in memory performance between the four familiar face categories (Figure 2b), as determined first by a 2×2 repeated-measures ANOVA with expression (EXP; smiling or angry) and feedback (FDB; won or lost) as factors, $F(15) < .711$, $p > .412$, and further confirmed by planned paired t -tests, $t(15) < 1.169$, $p > .261$.

To examine whether there were significant differences in the memory rate for some faces, we performed an additional item-based analysis that showed consistent effects for only 4 out of the 16 different face identities, with one male and one female being remembered significantly better than average, but one male and one female being remembered significantly worse than average (more “hits” and more “misses” relative to the mean for all faces, one-tailed t -test, $p < .05$). However, more importantly, there were no significant differences as a function of the feedback category (see Figure 2b), because face identities were assigned to the four different feedback conditions in a counterbalanced order, so as to

cancel out any visual memory effects related to a few specific faces overall.

After scanning, subjects also performed an explicit “sorting task” in which they had to classify old faces into the corresponding feedback categories of the encoding phase. There was no significant difference in accuracy rates between the four familiar face categories, as first determined by a 2×2 repeated-measures ANOVA with EXP (smiling or angry) and FDB (won or lost) as factors, $F(15) < .952$, $p > .345$, and further confirmed by planned paired t -tests, all $t(15) < 1.221$, $p > .241$. These results indicate that participants did not explicitly recall the exact feedback category in which old faces were seen during the encoding phase. Finally, subjects were also asked to make likeability ratings on old faces after scanning (using a five-point scale), but again there were no significant differences between the previous context conditions, as determined by a 2×2 repeated measures ANOVA with EXP (smiling or angry) and FDB (won or lost) as

factors, $F(15) < 1.297$, $p > .273$, and by paired t -tests, all $t(15) < 1.499$, $p > .155$.

Eye-tracking data

During encoding, there were no significant differences between the number and duration of fixations over the face in the four different social contexts, 2×2 repeated-measures ANOVA: all $F < 2.56$, $p > .14$; paired comparisons: all $t < 1.93$, $p > .08$. During recognition, there was a significant main effect of feedback on fixation duration, lost > won: $F = 7.42$, $p = .02$, but no significant main effect of expression and no interaction, all $F < 3.86$, $p > .075$. To ensure that increased fixation to faces did not modulate BOLD signal during recognition, we correlated the fixation times with beta values extracted from several ROIs identified in subsequent fMRI analysis (see below), but this revealed no significant effect, except for a positive correlation between responses to AF-L faces in left FG and fixation times, $r = .607$, $p = .036$. However,

TABLE 1

SPM results showing brain areas activated by familiar faces during the recognition test phase (main contrast “old” versus “new” faces, for correct answers only)

Voxels	T value	Z value	p value	Old correct > new correct			Region	BA
				x	y	z (mm)		
88	6.87	4.55	<.001	-54	24	33	Dorsal PFC left	44
48	4.22	3.38	.001	-42	45	0	Lateral PFC left	45
11	4.13	3.32	.001	51	33	30	Dorsal PFC right	45
21	4.02	3.26	.001	42	9	36	Lateral PFC right	44
12	4.14	3.33	.001	15	3	-18	Amygdala right	
6	3.67	3.05	.001	-15	-6	-15	Amygdala left	
171	13.36	6.11	<.001	-33	21	0	Anterior insula left	
54	6.68	4.48	<.001	57	-45	-3	Posterior STS right	21
8	3.77	3.11	.001	-51	-18	-21	Mid STG left	20
29	4.93	3.75	<.001	48	-27	-15	Mid STG right	20
62	4.67	3.62	<.001	33	18	-27	Temporal pole right	38
19	4.26	3.4	<.001	18	-69	30	Precuneus right	18/23
14	4.24	3.3	<.001	0	-36	27	Posterior cingulate	23
169	6.1	4.26	<.001	-51	-63	-18	Lateral fusiform gyrus left	37
21	4.77	3.66	<.001	-27	-66	33	Occipital cortex left	19
33	4.57	3.56	<.001	45	-75	-33	Occipital cortex right	19
35	5.31	3.92	<.001	33	-69	57	Parietal superior cortex right	7
13	4.24	3.39	<.001	9	-42	69	Paracentral cortex right	4/5
31	5.11	3.83	<.001	-42	-36	42	Parietal inferior cortex left	40
134	4.56	3.56	<.001	-48	-42	57	Parietal inferior cortex left	40
15	3.9	3.19	.001	48	-39	42	Supramarginal gyrus right	40
14	4.31	3.43	<.001	36	-51	39	Angular gyrus right	40
8	5.74	4.11	<.001	6	-84	-30	Cerebellum right	
5	4.89	3.73	<.001	33	-69	-39	Cerebellum right	
92	5.32	3.93	<.001	-6	-18	-3	Thalamus left	

Notes: Coordinates are given in MNI space. Activation sites were verified using the average anatomical MRI image of our 16 subjects. BA = Brodmann’s area, PFC = prefrontal cortex, STS = posterior superior temporal sulcus, STG = superior temporal gyrus.

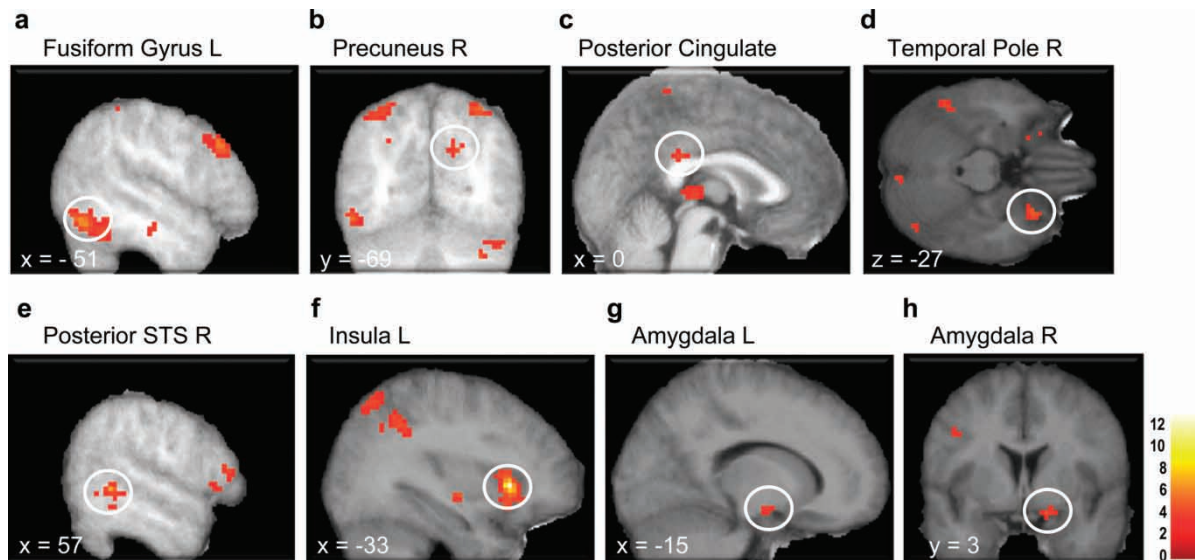


Figure 3. Main effect of face familiarity. Statistical parametric map for the contrast of “old faces” vs. “new faces” (correct answers only), showing an extensive network of brain areas responding to face familiarity. (a) Left fusiform gyrus ($xyz = -51 -63 -18$, $z = 4.26$, $p < .001$). (b) Right precuneus ($xyz = 18 -69 30$, $z = 3.40$, $p < .001$). (c) Posterior cingulate ($xyz = 0 -36 27$, $z = 3.38$, $p < .001$). (d) Right temporal pole ($xyz = 33 18 -27$, $z = 3.62$, $p < .001$). (e) Right superior temporal sulcus ($xyz = 57 -45 -3$, $z = 4.48$, $p < .001$). (f) Left insula ($xyz = -33 21 0$, $z = 6.11$, $p < .001$). (g) Left amygdala ($xyz = -15 -6 -15$, $z = 3.05$, $p < .001$). (h) Right amygdala ($xyz = 15 3 -18$, $z = 3.33$, $p < .001$). All p -values uncorrected, whole brain analysis.

this correlation cannot account for the fMRI results obtained in other conditions. Overall, these eye-tracking data therefore indicate that familiarity effects reported below cannot be explained by the visual inspection time for different faces, during either encoding or recognition.

fMRI results

We first identified brain regions involved in successful recognition of familiar faces by computing the contrast between all “old” faces vs. “new” faces, using correct answers only. This contrast allowed us to test for implicit and explicit memory processes together (unlike a comparison of hits vs. misses for “old” faces that would primarily test for explicit memory and confidence). Our analysis revealed a network of activated regions (see Table 1) including left FG, right posterior STS, right temporal pole (TP), as well as several limbic areas in left insula (INS), bilateral amygdala (AMY), and PCC (Figure 3). These areas were then used to define ROIs for further analysis (see below).

We also computed the inverse contrast (new faces vs. old faces), again for correct answers only, but this showed no significant activations at the same threshold.

Next, we specifically tested whether increases in activity for successful recognition of familiar faces would be modulated by the different context conditions in which faces were previously seen during the encoding phase. Our main goal was to determine any difference in brain responses to faces that were seen as “friends” or “allies” in the past (congruent feedback context, SF-W and AF-L), as opposed to faces that were previously seen as “foes” (incongruent feedback context, SF-L and AF-W). This comparison was performed in two steps, as follows.

In a first step, we extracted parameter estimates of activity (beta values) from ROIs that showed general increases to familiar faces, as determined by the main contrast of all “old” vs. “new” faces (correct answers only, see above and Table 1). We then split data from the familiar “old” face category into the four conditions of previous encounter: SF-W, AF-L, HF-L, and AF-W, respectively. Repeated-measure ANOVAs were performed on these parameter estimates to test whether activity in the ROIs was modulated by previous emotional expression (smiling or angry), previous feedback type (winning or losing), or any interaction between these two factors (which specifically determined the perceived social context of faces due to expression congruency/incongruency), in addition to direct

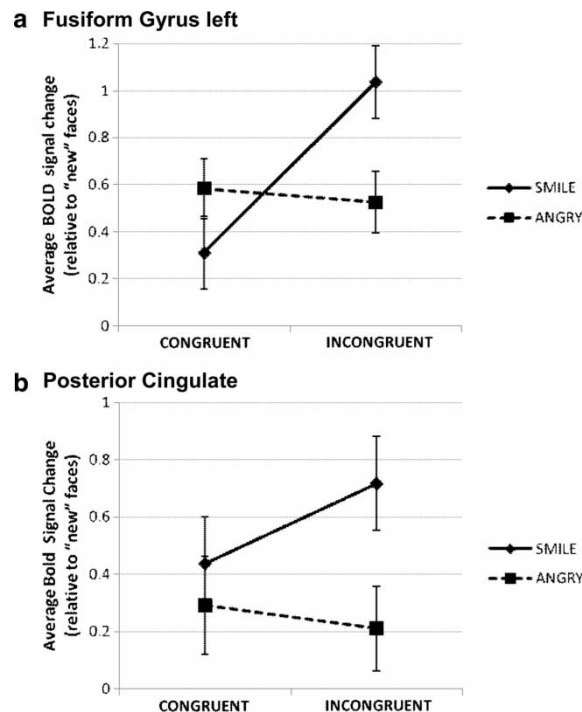


Figure 4. Activation of left fusiform gyrus and posterior cingulate across the familiar face conditions. (a) Parameter estimates (beta values) extracted from the left fusiform gyrus (contrast “old faces vs. new faces”), averaged across voxels and participants, showing a significant interaction between expression and feedback type, with the greatest response to faces previously seen with a smiling expression on LOST trials. (b) Parameter estimates (beta values) extracted from the posterior cingulate (same contrast), showing greater responses to faces previously seen with a smiling than angry expression. The average parameter estimates are shown relative to activation for the “new”/unfamiliar faces (by subtraction). All values are displayed with ± 1 SEM.

t-test comparisons to confirm significant increases for each of the “old” face conditions relative to “new” faces.

These analyses revealed that only three of these ROIs were modulated by the social significance of previous encounters. The left FG showed a significant main effect of previous feedback type, lost > won; $F(1, 15) = 5.97$, $p = .027$, and, more importantly, a significant interaction between feedback type and expression, $F(1, 15) = 5.05$, $p = .04$. The latter was due to the fact that faces of “foes” previously seen with a smile on lost trials (SF-L) activated the FG more than all other feedback conditions (Figure 4a). In addition, the PCC showed a marginally significant effect of previous expression, smiling > angry; $F(1, 15) = 4.34$, $p = .055$, but no effect of feedback, $F(1, 15) = .87$, $p = .36$, and no interaction, $F(1, 15) =$

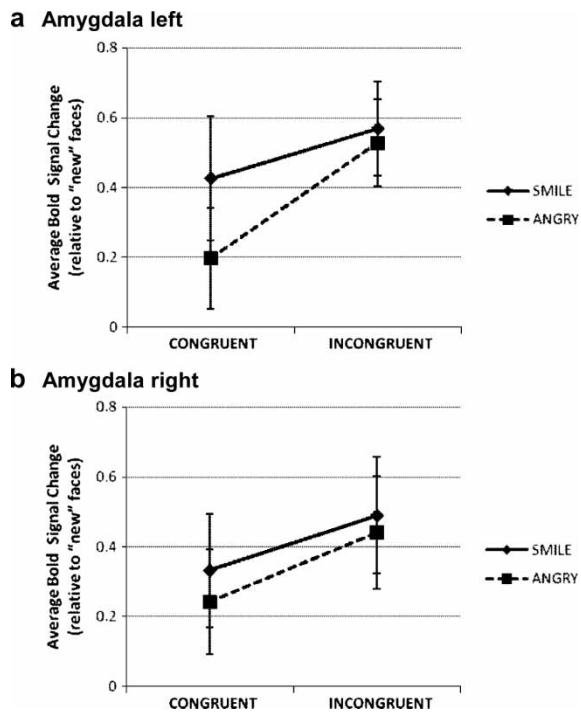


Figure 5. Activation of bilateral amygdala across the familiar face conditions. (a) Parameter estimates (beta values) extracted from the left amygdala (contrast “old faces vs. new faces”), averaged across voxels and participants, showing a greater response to faces previously seen in the two incongruent social context (smiling on LOST trials or angry on WIN trials). (b) Parameter estimates (beta values) extracted from the right amygdala showed the same pattern as the left amygdala. The average parameter estimates are shown relative to activation for the “new”/unfamiliar faces (by subtraction). All values are displayed with ± 1 SEM.

.34, $p = .57$ (Figure 4b). Finally, bilateral amygdala showed no significant effect in ANOVAs, but a pattern of selective increases for the two conditions of “foes” (SF-L and AF-W) relative to new (unfamiliar) faces, *left AMY*: SF-L > new, $t(15) = 3.42$, $p = .004$; AF-W > new, $t(15) = 3.22$, $p = .006$; *right AMY*: SF-L > new, $t(15) = 2.58$, $p = .021$; AF-W > new, $t(15) = 2.74$, $p = .015$; all paired *t*-tests (see Figure 5). By contrast, the amygdala response to the two conditions of “allied” faces (SF-W and AF-L) did not significantly differ from responses to new faces (all paired *t*-tests < 1.92, $p > .075$, on both sides). All other ROIs (including right STS, right temporal pole, and left insula) showed a general increase for the old/familiar face conditions as compared with the new/unfamiliar faces, but no differential effect of previous social context.

In a second step, we computed specific SPM contrasts on whole-brain data to directly compare

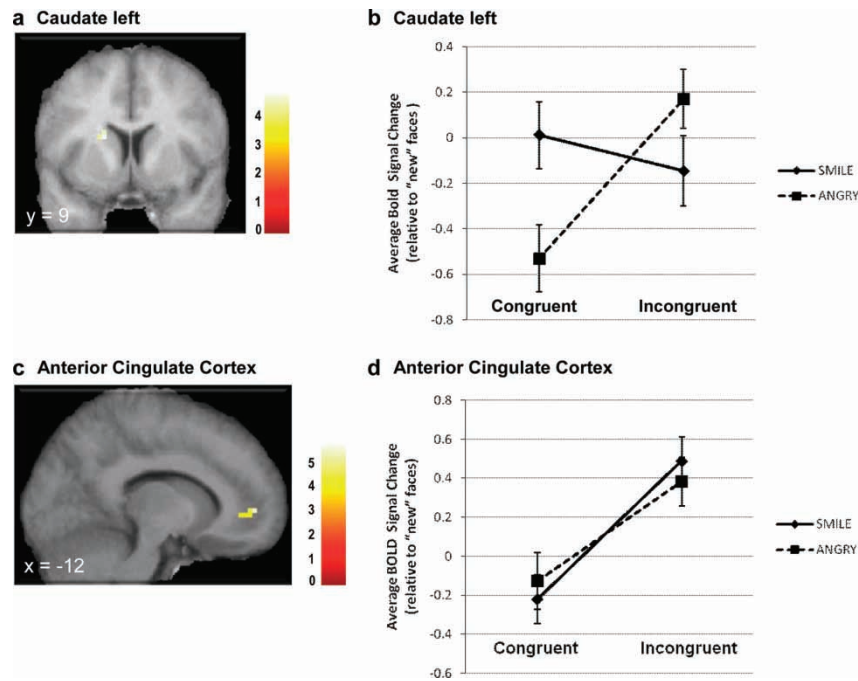


Figure 6. Activation to familiar faces previously seen in different feedback contexts. (a) Statistical parametric map for the contrast between faces seen on “won vs. lost” trials, showing increases in the left caudate nucleus ($xyz = -18\ 9\ 15$; T -value = 4.93; $p < .001$; 8 voxels). (b) Parameter estimates (beta values) extracted from the left caudate cluster, averaged across voxels and participants. (c) Statistical parametric map for the contrast between faces seen with “incongruent vs. congruent” expressions, showing increases in the anterior cingulate cortex ($xyz = -12\ 51\ 0$; T -value = 3.86; $p < .001$; 23 voxels). (d) Parameter estimates (beta values) extracted from the anterior cingulate cluster, averaged across voxels and participants. The average parameter estimates are shown relative to activation for the “new”/unfamiliar faces (by subtraction). All values are displayed with ± 1 SEM.

the different conditions of old faces with each other. We tested for any main effect of previous emotional expression (EXP; smiling vs. angry), previous feedback type (FDB; winning vs. losing), and previous social significance of the face context (congruent vs. incongruent expression, yielding the perception of “allied” or “friend” faces vs. “foes”, respectively). This analysis revealed two additional regions that were modulated by past impressions, including the left caudate nucleus (Figure 6a) when comparing faces seen on “winning” vs. “losing” trials, and the ACC (Figure 6c) when comparing faces with incongruent (“foes”) versus congruent (“friends”) expressions.

For the left caudate nucleus (Figure 6b), subsequent pairwise comparisons showed that this increased activation to faces previously seen on WON trials was essentially due to a significant difference for angry faces associated with the incongruent feedback condition (i.e. “foes”), AF-W vs. AF-L: $t(15) = 3.35$, $p = .004$, paired t -test; but there was no difference between WON vs. LOST trials for smiling faces, SF-W vs. SF-L: $t(15) = .77$, $p = .45$. In contrast, for the rostral ACC (Figure 6d), pairwise comparisons showed that there were

significant increases for both angry and smiling faces when previously associated with an incongruent feedback context (AF-W and SF-L, “foes”), with a greater activation to angry faces from WON relative to LOST trials, AF-W vs. AF-L: $t(15) = 2.19$, $p = .045$; paired t -test, but conversely, greater activation to smiling faces from LOST relative to WON trials, SF-L vs. SF-W: $t(15) = 3.65$, $p = .002$.

DISCUSSION

The current study investigated the neural substrates of social memory for positive and negative person traits, which were learned by exposure to faces associated with different social contextual cues. Face identities seen by the participants were those of strangers who became familiar through a pseudo-interactive social game, allowing us to avoid any knowledge-related or intrinsic emotional differences prior to exposure. After this first encoding phase, face recognition was probed while brain activity was measured by fMRI, and neural responses were compared as a function of

the different conditions in which the faces were seen during initial exposure. Importantly, all faces were presented with a neutral expression during the recognition phase, whereas they were repeatedly seen with a distinct emotional expression and paired with a distinct feedback context during initial exposure. Thus, by orthogonally manipulating emotional expression (smiling vs. angry) and feedback context (positive vs. negative), we could induce a different social meaning for each face identity even when they shared the same expression (e.g. a smile praising success or mocking a failure). This allowed us to distinguish between subsequent effects in memory due to the past emotional expression alone (which might result from relatively automatic perceptual analysis), and subsequent effects due to systematic association with positive or negative context, or any interactions between these two factors that actually determined the perceived social significance of alliance or hostility for the viewer.

Our results reveal that familiarity of the faces induced by previous encounters in the game task produced increased BOLD responses in a widespread network of brain regions previously reported to be involved in familiar face recognition (Gobbini & Haxby, 2007), including the left fusiform gyrus, right posterior superior temporal sulcus, right precuneus, posterior cingulate, right temporal pole, left insula, and bilateral amygdala. More critically, we found that brain regions within both the “core system” of face perception and “extended system” of person recognition (Gobbini & Haxby, 2007) were generally modulated by familiarity alone, with no or little differences as a function of the past social significance of individual faces (i.e., the four different feedback conditions). Only the left fusiform gyrus showed relatively greater increases for faces perceived as hostile in the past (SF-L condition), while posterior cingulate cortex showed relatively greater increases for faces seen with a smile in the past (SF-W and SF-L). In contrast, subsequent effects of social context were found in brain regions associated with motivational processes and behavioral monitoring, such as the amygdala, caudate nucleus, and anterior cingulate, which showed strongest or even selective responses to faces previously perceived as “foes” (i.e., smiling when the participant failed and frowning when the participant succeeded in the task, SF-L and AF-W conditions, respectively). Although ACC and caudate have traditionally not been integrated in human brain networks for person recognition

(Adolphs, 2003; Gobbini & Haxby, 2007), our new data converge with other recent findings in humans (de Quervain et al., 2004; Somerville et al., 2006) and monkeys (Rushworth, Behrens, Rudebeck, & Walton, 2007) to suggest that these regions may play a crucial role in social cognition and perhaps more generally in affect regulation.

Note that these modulations were not due to previous exposure to different expressions or different feedback types alone; rather, such differences were specifically induced by the interaction between facial expression and feedback context which conveyed a distinctive social meaning to previously seen faces. Likewise, different brain responses to faces encountered in different social contexts could not be explained by differences in eye movements, as online eye-tracking measures did not show any difference between the four feedback conditions – a factor not systematically taken into account in previous studies of familiar face recognition (Gobbini & Haxby, 2007). These new imaging data are consistent with the notion that person impressions can readily be formed even during brief social encounters with people (Bar et al., 2006; Todorov et al., 2007; Willis & Todorov, 2006), and then efficiently transfer to a new view of the same person (despite changes in appearance and expression). Furthermore, these findings also confirm an important role of situational context for inducing such enduring person impressions (Gilbert & Malone, 1995).

Below we will discuss activation patterns for the most relevant brain regions.

Modulations by familiarity in the “core” and “extended” face processing systems

Among regions showing greater activation to previously seen faces than to new faces, many were essentially sensitive to their past familiarity irrespective of specific social context and induced person impression, such as right posterior STS, right precuneus, right temporal pole, and left insula. These regions are likely to be implicated in various aspects of person knowledge (Gobbini & Haxby, 2007). In particular, STS has previously been found to adapt to specific combinations of face identity and expression (Winston, Henson, Fine-Goulden, & Dolan, 2004), and to activate when processing cues that imply intentional action (Gallagher & Frith, 2003) or mentalizing about intentions of others (Frith & Frith, 2003).

Moreover, STS activity has already been related to implicit social judgments during face processing tasks (Gobbini et al., 2004; Leibenluft et al., 2004; Singer et al., 2004; Todorov et al., 2007). Our findings therefore suggest that prior experience with faces in a social context as implemented here during initial exposure may then enhance the perception of intentionality and mental states for these faces, as compared to new faces without any previous personal history. However, such activity was not significantly modulated by the content of mental state attribution, since no statistical difference was observed as a function of the previous context.

Similarly, the right precuneus and right temporal pole were activated by face familiarity but not by their previous expression, feedback, or contextual associations. Precuneus does not only activate to familiar faces in general (Gobbini et al., 2004; Leibenluft et al., 2004), but also during mental imagery (Fletcher et al., 1995; Ishai, Haxby, & Ungerleider, 2000), and long-term and episodic memory retrieval (Burgess, Maguire, Spiers, & O'Keefe, 2001; Fletcher et al., 1995). Likewise, anterior temporal regions may respond to various familiar stimuli, including faces (Gorno-Tempini et al., 2000), names (Grabowski et al., 2001), or landscapes (Nakamura et al., 2000). Moreover, lesions in precuneus and/or temporal poles may impair recognition and access to semantic information about people (see e.g. Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Vuilleumier, Mohr, Valenza, Wetzell, & Landis, 2003). These findings suggest that both the precuneus and temporal poles might contribute to familiar face recognition by subserving general processes associated with mental imagery and retrieval of episodic representations in memory, but do not code for more specific social or affective representations influenced by previous context in which faces are encountered.

The left insula was another region modulated by familiarity irrespective of past person impression. Insula responses to familiar faces have already been observed not only for one's own versus unrelated children (Leibenluft et al., 2004), but also for faces previously experienced as cooperators in an economic exchange (Singer et al., 2004) or perceived as aggressive and disgusting based on fictive scenarios (Todorov et al., 2007). These findings suggest that insula responses may not differentiate between specific affective values (i.e., positive or negative) associated with familiar faces,

but represent a more general affective response related to increased autonomic arousal (Critchley, Melmed, Featherstone, Mathias, & Dolan, 2002) which might be elicited by the recognition of familiarity and/or the retrieval of information about known persons and past encounters with them.

Modulations by person impression in the face processing system

In contrast to the above regions, left fusiform cortex not only showed greater responses to successfully recognized familiar faces, as compared with new faces, but this effect was enhanced when faces were previously paired with negative feedback (lose vs. win), and more specifically when this was combined with a smiling expression that induced negative social impressions of hostility ("foes", SF-L condition, cf. Figure 4a). These results suggest that face processing in this condition might have been modulated by top-down influences from other brain areas associated with social or affective memory, based on negative person impressions elicited by past experience. Similar modulations of face processing in visual cortex by familiarity or affective value have already been reported in other studies in humans (Gobbini & Haxby, 2007; Singer et al., 2004) and single-cell recordings in monkeys (Sugase, Yamane, Ueno, & Kawano, 1999). In our experiment, this modulation of fusiform activation appeared particularly driven by the combination of expression and feedback context, rather than by any of these factors alone, suggesting that it could not simply be explained by different visual information perceived during encoding. Such top-down modulation of fusiform activity might possibly originate from amygdala, since this region is critically involved in affective and social processing, as well as in learning (Adolphs, 2003; Dolan, 2002), and thought to modulate fusiform activation to emotional faces (Vuilleumier et al., 2001). In our study, the amygdala and ACC were also activated by faces associated with negative social context ("foes", see Figures 5 and 6 and below), and both of these regions might therefore contribute to modulatory signals affecting fusiform activity (directly or indirectly).

Finally, the PCC was another region within the person recognition network that showed general increases to all familiar faces (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005; Shah

et al., 2001), relative to new faces, as well as a modulation by the past social context. PCC activity was significantly enhanced when faces were previously seen with a smiling (vs. angry) expression. Other studies reported activation of PCC by autobiographical memory retrieval concerning familiar people (Maddock, Garrett, & Buonocore, 2001; Shah et al., 2001), judging emotional valence or content for words and other stimuli (Maddock, Garrett, & Buonocore, 2003), and judging the morality of one's own or someone else's behavior (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Our results support the notion that PCC might be involved not only in appraising the affective significance of a stimulus (for oneself or others), but also in modulating memory retrieval in relation to affectively relevant stimuli. However, unlike previous data suggesting that PCC activates independent of emotional valence (e.g., for both pleasant and unpleasant arousing words; (Maddock et al., 2001), our finding of greater activation by previously smiling faces than previously angry faces reveals a bias for positive emotional values associated with people, which might account for reliable responses to personally familiar faces in previous imaging studies (cf. Pourtois et al., 2005; Shah et al., 2001).

Modulations in emotion and prefrontal systems

A major finding of our study was that bilateral amygdala, left caudate nucleus, and ACC showed increases to faces that elicited negative impressions of hostility during encoding (SF-L and AF-W conditions), reflecting that a distinctive social value (i.e. "foes") was retained from prior context and modulated activity in these regions during a subsequent encounter with the same identities.

The amygdala was found to activate bilaterally for familiar faces, but more specifically in the two "foe" conditions (see Figure 5). This finding might help reconcile apparent discrepancies in the literature on familiar person recognition. On one hand, several studies reported weaker amygdala response to known vs. unknown faces (Dubois et al., 1999), personally familiar vs. famous faces and strangers (Gobbini et al., 2004), or a lover vs. a stranger face (Bartels & Zeki, 2004), which was interpreted as lower vigilance or threat when encountering a known person or reduced need to assess social integrity.

This interpretation also accords with greater amygdala activation by the perception of untrustworthy faces (Winston et al., 2002) or faces from racial outgroups (Phelps, Cannistraci, & Cunningham, 2003), in keeping with a more general role for appraising potential threats (Breiter et al., 1996; Morris et al., 1996). On the other hand, some studies reported increased amygdala activation to pictures of one's own children vs. unknown or familiar but unrelated children (Leibenluft et al., 2004), as well as to cooperators relative to defectors in an economic game (Singer et al., 2004), which was attributed to increased vigilance, protectiveness, or reward expectation for close social relationships. Since in our task amygdala response was stronger to familiar faces than to new faces, but also stronger to familiar faces associated with incongruent social feedback than to those associated with congruent feedback, our results add support to the notion that the human amygdala might primarily encode self-relevant information (Sander, Grafman, & Zalla, 2003) and act to increase vigilance in conditions of ambiguity (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003; Kim et al., 2004) or uncertainty (Hsu, Bhatt, Adolphs, Tranel, & Camerer, 2005). This notion would fit well with our data, because amygdala activity did not appear to track the valence of past expression or past feedback *per se*, but rather the personal relevance of past experiences with faces.

The left caudate nucleus also exhibited an activation pattern that selectively discriminated faces of "foes" previously seen with an angry expression on WON as opposed to LOST trials (see Figure 6a, b). As caudate activity has previously been related to the representation of rewards in social and nonsocial domains (Kampe, Frith, Dolan, & Frith, 2001), this pattern might reflect a rewarding value that is experienced when re-encountering faces of hostile partners who were outplayed during the social game. These findings converge with data from previous studies investigating brain responses related to altruistic punishment (de Quervain et al., 2004) and processing of social hierarchy in humans (Zink et al., 2008). In these two studies, the ventral striatum (including the caudate nucleus) was more activated when subjects could effectively punish defectors in an economic exchange, or moved to a higher hierarchical position based on their performance against a previously superior player. Our data thus provide new evidence that "superiority" information (i.e., winning

against the conflicting goals of a virtual opponent in our pseudo-game) is an important component of social memory traces formed after exposure to faces, which may contribute to define the social status of a previously encountered person, and that such information is at least partly dependent on activity of reward circuits in caudate nucleus.

Finally, of particular interest, in addition to regions typically involved in familiar face recognition, we found that activity in the rostral ACC was significantly modulated by the prior social context of faces, and exhibited a selective increase to faces of “foes” from the two incongruent feedback conditions. This region accords well with a functional segregation recently proposed for the medial frontal cortex (MPFC; Amodio & Frith, 2006), including an anterior region (aMPFC) that may be particularly important to access emotional and social knowledge concerning the self or other persons, such as during mentalizing (Frith & Frith, 1999, 2003) or introspection processes (Mitchell, Banaji, & Macrae, 2005a, 2005b). The rostral anterior ACC identified in our study overlapped with this aMPFC subregion (Amodio & Frith, 2006), and with increases previously observed when thinking about the mental states of other people or animals (Kelley et al., 2002; Mitchell et al., 2005a, 2005b; Schmitz, Kawahara-Baccus, & Johnson, 2004). Another study (Somerville et al., 2006) reported that a similar region ($xyz = -6\ 45\ 3$) was modulated by the perception of social rejection during a pseudo-interactive paradigm in which participants read the judgment of other virtual participants on them (e.g. “This person said she did not like you”). The aMPFC might therefore play an important role in appraisal processes that serve to predict social and/or emotional outcomes, not only for ongoing social events but also based on memory traces from past encounters.

Because most of the abovementioned areas showed the highest levels of activation for familiar faces previously encountered in an incongruent social context, our results could potentially also be interpreted as a simple incongruency/congruency effect on memory. However, an incongruency effect would predict similar influences for both types of incongruent faces irrespective of their previous expression (smiling or angry), whereas our results indicated a clear difference as a function of expression, in agreement with a role for more specific factors related to affective/social perception during encoding. Moreover, all brain

areas showing these differential effects have previously been related to social processing, further suggesting that incongruency *per se* is unlikely to explain the present activation patterns.

Implicit vs. explicit memories of person impressions

Strikingly, differential brain responses to faces were observed depending on prior social context although our behavioral “sorting” test post-fMRI revealed that participants were unable to explicitly classify familiar faces into the correct condition of past context in which they had been seen. Moreover, we found no significant differences in memory performance between the four distinct feedback conditions (cf. Figure 2b). Although it is not possible to draw conclusions from the absence of significance, these data suggest that the differential responses reflecting past person impressions were likely to be implicit or unconscious, which would corroborate the notion that person impressions may be formed automatically without awareness (Uleman et al., 2005), and converge with other fMRI data showing implicit biases in brain responses to specific categories of faces (e.g., outgroup race effects; Phelps et al., 2003) in the absence of explicit acknowledgment by the participant. Our data therefore also add support to behavioral findings that the retrieval of affective person knowledge may arise without any explicit memory for the behaviors that triggered these affective inferences. Accordingly, a recent fMRI study (Todorov & Uleman, 2002) reported that activation of STS by affective person knowledge during face recognition was dependent on explicit memory for past behaviors, whereas activation in insula was implicit (Todorov et al., 2007). In our study, however, both regions were activated by familiar faces despite a lack of subsequent explicit memory for the exact context of past encounters; but our participants were engaged in a memory task (old/new recognition) which did not allow us to directly address the exact nature of memory traces. Further research is necessary to delineate a more specific contribution of different brain areas to distinct components of explicit and implicit representations of person impressions in memory, by including additional behavioral measures such as skin conductance responses.

CONCLUSIONS

Our study shows that a wide network of brain areas involved in face recognition is modulated by familiarity of faces derived from past encounters in different contexts and with different expressions, including left fusiform gyrus, right posterior superior temporal sulcus, left insula, right temporal pole, right precuneus, and posterior cingulate cortex. Furthermore, the activation of some of these regions to new encounters of the same face identities is modulated by the positive (friendly) or negative (hostile) nature of past social context, providing a neural substrate for enduring person impression that can influence emotion and behavior during subsequent presentation with the same face identities. Critically, we provide new evidence that the ventral anterior cingulate cortex, together with amygdala, are key components in neural systems activated by negative person impressions, possibly engaged by retrieval of affective person knowledge so as to adjust emotional responses to faces associated with conflict behavior during previous social encounters. In addition, we show that caudate nucleus may code for information concerning social superiority and selectively activates to faces of outplayed opponents, possibly serving to guide behavior based on the relative social status of previously encountered people. These results not only contribute to better delineating brain systems underlying social cognition skills in humans, but also help to understand neurobiological factors involved in the perception and reaction to social conflicts and aggression behaviors.

Manuscript received 13 October 2008
 Manuscript accepted 31 March 2009
 First published online day/month/year

REFERENCES

- Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*(3), 165–178.
- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*(6684), 470–474.
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277.
- Armony, J. L., Chochol, C., Fecteau, S., & Belin, P. (2007). Laugh (or cry) and you will be remembered—Influence of emotional expression on memory for vocalizations. *Psychological Science*, *18*(12), 1027–1029.
- Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278.
- Bargh, J. A. (1994). The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage*, *21*(3), 1155–1166.
- Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, *17*(5), 875–887.
- Burgess, N., Maguire, E. A., Spiers, H. J., & O’Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *NeuroImage*, *14*(2), 439–453.
- Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience*, *3*(11), 1077–1078.
- Critchley, H. D., Melmed, R. N., Featherstone, E., Mathias, C. J., & Dolan, R. J. (2002). Volitional control of autonomic arousal: A functional magnetic resonance study. *NeuroImage*, *16*(4), 909–919.
- Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, *380*(6574), 499–505.
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1–2), 179–229.
- D’Argembeau, A., & Van der Linden, M. (2007). Facial expressions of emotion influence memory for facial identity in an automatic way. *Emotion*, *7*(3), 507–515.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*(11), 1611–1618.
- de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schelhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*(5688), 1254–1258.
- Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194.
- Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., et al. (1999). Effect of familiarity on the processing of human faces. *NeuroImage*, *9*(3), 278–289.
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, *302*(5643), 290–292.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519.
- Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S. J., & Dolan, R. J. (1995). The

- mind's eye—Precuneus activation in memory-related imagery. *NeuroImage*, 2(3), 195–200.
- Frith, C. D., & Frith, U. (1999). Cognitive psychology—Interacting minds—A biological basis. *Science*, 286(5445), 1692–1695.
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 358(1431), 459–473.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, 7(2), 77–83.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117(1), 21–38.
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1), 32–41.
- Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage*, 22(4), 1628–1635.
- Gorno-Tempini, M. L., Price, C. J., Josephs, O., Vandenberghe, R., Cappa, S. F., Kapur, N., et al. (2000). The neural systems sustaining face and proper-name processing (vol 121, pg 2103, 1998). *Brain*, 123, 419–419.
- Grabowski, T. J., Damasio, H., Tranel, D., Ponto, L. L. B., Hichwa, R. D., & Damasio, A. R. (2001). A role for left temporal pole in the retrieval of words for unique entities. *Human Brain Mapping*, 13(4), 199–212.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108.
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science*, 310(5754), 1680–1683.
- Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2000). Distributed neural systems for the generation of visual images. *Journal of Cognitive Neuroscience*, 12(1), 70–70.
- Kampe, K. K. W., Frith, C. D., Dolan, R. J., & Frith, U. (2001). Reward value of attractiveness and gaze—Making eye contact enhances the appeal of a pleasing face, irrespective of gender. *Nature*, 413(6856), 589–589.
- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, 14(5), 785–794.
- Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport*, 14(18), 2317–2322.
- Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience*, 16(10), 1730–1745.
- Leibenluft, E., Gobbini, M. I., Harrison, T., & Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their children and other children. *Biological Psychiatry*, 56(4), 225–232.
- Maddock, R. F., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: The posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, 104(3), 667–676.
- Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping*, 18(1), 30–41.
- Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, 17(2), 119–128.
- Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005a). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *NeuroImage*, 28(4), 757–762.
- Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005b). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, 17(8), 1306–1315.
- Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, 383(6603), 812–815.
- Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing—A PET study. *Brain*, 123, 1903–1912.
- Oda, R. (1997). Biased face recognition in the Prisoner's Dilemma game. *Evolution and Human Behavior*, 18(5), 309–315.
- Phelps, E. A., Cannistraci, C. J., & Cunningham, W. A. (2003). Intact performance on an indirect measure of race bias following amygdala damage. *Neuropsychologia*, 41(2), 203–208.
- Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F. O., & Vuilleumier, P. (2005). View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: An event-related fMRI study. *NeuroImage*, 24(4), 1214–1224.
- Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences*, 11(4), 168–176.
- Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences*, 14(4), 303–316.
- Schmitz, T. W., Kawahara-Baccus, T. N., & Johnson, S. C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *NeuroImage*, 22(2), 941–947.
- Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H. J., et al. (2001). The neural correlates of person familiarity—A functional magnetic resonance imaging study with clinical implications. *Brain*, 124, 804–815.
- Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, 41(4), 653–662.

- Somerville, L. H., Heatherton, T. F., & Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nature Neuroscience*, *9*(8), 1007–1008.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*(6747), 869–873.
- Takahashi, T. (2005). Social memory, social stress, and economic behaviors. *Brain Research Bulletin*, *67*(5), 398–402.
- Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*(1), 163–173.
- Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051–1065.
- Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*(6), 549–562.
- Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.
- Trope, Y., & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology*, *73*(4), 662–674.
- Uleman, J. S., Blader, S., & Todorov, A. (2005). Implicit impressions. In J. S. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious*. New York: Oxford University Press.
- Vrtička, P., Andersson, F., Grandjean, D., Sander, D., & Vuilleumier, P. (2008). Individual attachment style modulates human amygdala and striatum activation during social appraisal. *PLoS ONE*, *3*(8), 2868.
- Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, *30*(3), 829–841.
- Vuilleumier, P., Mohr, C., Valenza, N., Wetzell, C., & Landis, T. (2003). Hyperfamiliarity for unknown faces after left lateral temporo-occipital venous infarction: A double dissociation with prosopagnosia. *Brain*, *126*, 889–907.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.
- Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, *92*(3), 1830–1839.
- Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–283.
- Zink, C. F., Tong, Y. X., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. *Neuron*, *85*(2), 273–283.