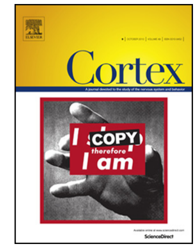
CORTEX 49 (2013) 2484–2493

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex




Research report

# Decoding the representation of learned social roles in the human brain

*Evelyn Eger* [a,b,c,*], *Laura Moretti* [d], *Stanislas Dehaene* [a,b,c,e] and *Angela Sirigu* [d]

[a] *Institut National de la Santé et de la Recherche Médicale, Unit 992, Gif/Yvette, France*
[b] *Commissariat à l'Energie Atomique, NeuroSpin, Gif/Yvette, France*
[c] *Université Paris-Sud, Orsay, France*
[d] *Centre National de la Recherche Scientifique, Unit 5229, Bron, France*
[e] *Collège de France, Paris, France*

ARTICLE INFO

*Article history:*
Received 7 May 2012
Reviewed 20 August 2012
Revised 9 November 2012
Accepted 6 February 2013
Action editor Alan Sanfey
Published online 20 February 2013

*Keywords:*
Functional magnetic resonance imaging
Multivariate decoding
Face processing
Social cognition
Learning

ABSTRACT

Humans as social beings are profoundly affected by exclusion. Short experiences with people differing in their degree of prosocial behaviour can induce reliable preferences for including partners, but the neural mechanisms of this learning remain unclear. Here, we asked participants to play a short social interaction game based on "cyber-ball" where one fictive partner included and another excluded the subject, thus defining social roles (includer – "good", excluder – "bad"). We then used multivariate pattern recognition on high-resolution functional magnetic resonance imaging (fMRI) data acquired before and after this game to test whether neural responses to the partners' and neutral control faces during a perceptual task reflect their learned social valence. Support vector classification scores revealed a learning-related increase in neural discrimination of social status in anterior insula and anterior cingulate regions, which was mainly driven by includer faces becoming distinguishable from excluder and control faces. Thus, face-evoked responses in anterior insula and anterior cingulate cortex contain fine-grained information shaped by prior social interactions that allow for categorisation of faces according to their learned social status. These lasting traces of social experience in cortical areas important for emotional and social processing could provide a substrate of how social inclusion shapes future behaviour and promotes cooperative interactions between individuals.



## 1. Introduction

The need for social affiliation is central to normal human existence. Therefore, the act of social exclusion (or ostracism) practised across human societies and cultures and even by some non-human primates, is usually perceived as a powerful and emotionally distressful signal, even under very artificial circumstances such as a computer game (Williams et al., 2000). Being totally or partially socially excluded can lead the affected individual to try to conform or re-establish social links with the group (Williams et al., 2000), and to develop emotional preferences for partners with a higher tendency to include them (Andari et al., 2010). Functional imaging studies recording brain activity during experiences of social exclusion

* *Corresponding author*. INSERM, U. 992, CEA/Neurospin, Batiment 145, Point Courrier 156, 91191 Gif/Yvette, France.
E-mail address: evelyn.eger@gmail.com (E. Eger).
0010-9452/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cortex.2013.02.008

have revealed that the "social pain" experienced under these conditions seems to share neural substrates with physical pain, i.e., activations in right anterior insula, anterior cingulate, and lateral prefrontal cortex (Eisenberger et al., 2003; Masten et al., 2009; Sebastian et al., 2011). Intracranial electrophysiological recordings have also shown effects on theta power in insula and subgenual anterior cingulate cortex during the experience of exclusion (Cristofori et al., 2012). However, all of these studies were restricted to measuring activity directly while subjects experienced the exclusion situation. The underlying neural mechanisms of how people develop preference or aversion for partners following social interactions still remain largely unclear. Here, we explore the brain correlates of such learned social categorisations, using a modified version of the cyber ball game involving fictive partners with different profiles (includer/excluder). Previous behavioural studies have shown that in such a situation, normal volunteers quickly start to favour partners that included rather than excluded them, as measured by the number of ball tosses sent to those partners and various behavioural ratings such as trust and preference (Andari et al., 2010).

We tested the hypothesis that activity in the brain structures mediating the emotional reaction during exclusion would also carry information reflecting the learnt social categories when the involved partners are subsequently encountered in a context lacking explicit social or affective connotations. We measured with fMRI brain responses to different facial identities before and after a very brief social interaction (ball tossing game) with different partners. The analysis methods we used were based on pattern recognition (support vector classification) which predict from the subject's brain activity the face's social meaning as previously experienced in the ball tossing game (includer = "good"; excluder = "bad"; control face = "neutral"). This approach, often applied in other domains of imaging neuroscience for example to study perceptual representations (Haynes and Rees, 2006; Norman et al., 2006), is applied here, to our knowledge for the first time, to decode learned social attributes. Multivariate pattern recognition methods allow to differentiate between experimental conditions on the basis of the information present in the full pattern of activity across voxels and are therefore sensitive to distributed effects that may remain undetected by conventional mass-univariate mapping procedures testing only for circumscribed activation in- or decreases. In the present context, this approach permits to detect subtle changes in distributed activity patterns in regions of interest (ROIs) relevant for social emotion that arise with learning and discriminate between faces as a function of their learned social category.

## 2. Methods

### 2.1. *Participants and fMRI acquisition*

15 healthy young volunteers (11 male and 4 female aged 24.1 ± 3.8 years) were included in this study which had been approved by the regional ethics committee of Hôpital de Bicêtre, France. Functional images were acquired on a 3 T magnetic resonance (MR) system (Siemens Tim Trio) with 12-channel head coil as T2* weighted echo-planar image (EPI) volumes with 1.5 mm in-plane resolution. 33 transverse slices covering occipital, temporal and frontal lobes up to (ventral parts of) anterior cingulate cortex were obtained in interleaved order (repetition time (TR) 3 sec, field of view (FOV) 192 mm, echo time (TE) 30 msec, flip angle 78°, slice thickness 2 mm).

### 2.2. *Stimuli and paradigm*

The experiment consisted of a pre-learning and a post-learning phase during which fMRI data were recorded, and which were separated by the learning phase (social interaction game) without fMRI acquisition, see Fig. 1A. Stimuli were back-projected onto a screen located at a distance of 1 m from the subjects eyes at the end of the scanner bore and viewed via a mirror attached to the head coil. In the two rounds of the cyber ball game (one with male, and one with female partners), two face images were displayed subtending 3.5 × 4.5° visual angle (VA) at ~7.5° in the left and right visual field alongside cartoon characters representing the other players (see Fig. 1B). The partner defined as excluder ("bad"), after a short unbiased period, very rarely (20% of tosses) sent the ball to the subject whereas the includer ("good") directed 50% of tosses each to the participant and the other partner. Thus, different from Andari et al. (2010), who in a three-partner version of the game used an over-including (80/20) "good" partner, here we defined the "good" partner by a 50/50 profile since we had behavioural evidence of the effectiveness of this two-partner version with normal subjects and without monetary rewards at the time of the planning of the current study, and we wanted to avoid situations of an overly one-sided exchange between participant and includer. Moreover, in previous work, over-inclusion did not yield a clear advantage over being equally included (Williams et al., 2000).

The pictures of faces used for this studies (three men, three women) were chosen from the NimStim database (Tottenham et al., 2009). Starting with a pool of 20 faces with neutral expression, each of these faces was rated for attractiveness by a group of 10 subjects (not involved in the fMRI experiment) on a seven-point scale ranging from 1 (extremely unattractive) to 7 (extremely attractive). Based on these scores, we selected six faces (three female, three males) whose attractiveness scores were in the average range (between 3 and 4). Assignment of individual face identities to the three experimental conditions ("good" *vs* "bad" faces in the game *vs* "neutral" control faces not appearing in the game) was close to counterbalanced across subjects (fully counterbalanced within 12 subjects). "Good" and "bad" partners were assigned to opposite sides across the two rounds of the games (e.g., "good" = left, "bad" = right in first round, "bad" = left, "good" = right in second round). The choice of an unrelated face that did not appear in the game as "neutral" face condition, while making the design slightly unbalanced, was a compromise motivated by the need to keep the cyber ball game period sufficiently short and maximise time for fMRI acquisitions.

During the pre-learning and post-learning phases preceeding and following the ball tossing game, subjects viewed mini-blocks (four presentations of the same face, 1 sec on, .5 sec off) of the six different facial identities (three male, three