Please cite this article in press as: Berlingeri M et al. *Guess who's coming to dinner*: Brain signatures of racially biased and politically correct behaviors. Neuroscience (2016), http://dx.doi.org/10.1016/j.neuroscience.2016.06.048

Neuroscience xxx (2016) xxx–xxx

GUESS WHO'S COMING TO DINNER: BRAIN SIGNATURES OF RACIALLY BIASED AND POLITICALLY CORRECT BEHAVIORS

M. BERLINGERI, [a,e*] **M. GALLUCCI,** [b,e] **L. DANELLI,** [b,e] **M. FORGIARINI,** [b] **M. SBERNA** [c] **AND E. PAULESU** [b,d,e*]

[a] *Department of Humanistic Studies (DISTUM), University of Urbino Carlo Bo, Urbino, Italy*

[b] *Psychology Department, University of Milano-Bicocca, Milano, Italy*

[c] *Neuroradiology Department, Niguarda Cà Granda Hospital, Milano, Italy*

[d] *IRCCS Galeazzi, fMRI Unit, Milano, Italy*

[e] *NeuroMi, Milan Center for Neuroscience, Italy*

Abstract—The ability to emphatically share feelings with those of someone in pain is affected by the racial difference between the target and the onlooker. A differential empathic activation for race (*DEAR effect*) in favor of in-group members has been documented in the brain pain matrix. However, we are also capable of unbiased responses that manifest *politically correct* behaviors toward people of a different race. To address the neurofunctional signatures underlying both the DEAR effect and the manifestation of politically correct behaviors, Caucasian participants performed an fMRI session in which videos were presented of either African or Caucasian actors touched by either a rubber eraser or a needle. Participants were instructed to empathize with the actors during the video presentation (stimulus phase) and to explicitly judge the pain level experienced by the actors (response phase). During the stimulus phase, we found a typical in-group-specific DEAR effect within the pain-matrix. This effect correlated with the level of implicit racial bias as measured by the IAT. On the other hand, during the response phase a significant out-group-specific DEAR effect emerged in the prefrontal cortices. This latter effect was coupled with a revealing behavioral pattern. That is, while the magnitude of the painful experience attributed to Caucasians and Africans was the same, our participants were significantly slower when judging the African's pain experience. We propose a model that logically integrates these *two contrasting forces at the neurobiological and behavioral level*. © **2016 Published by Elsevier Ltd on behalf of IBRO.**

Key words: empathy, fMRI, racial bias, embodied responses, controlled processes, politically correct responses.

*Corresponding authors. Address: Dipartimento di Studi Umanistici, DISTUM, Università degli Studi di Urbino Carlo Bo, Via Bramante, 17, 61029 Urbino (PU), Italy (M. Berlingeri). Dipartimento di Psicologia, Università degli studi di Milano-Bicocca, Piazza dell'ateneo nuovo 1, edificio U6, 20126 Milano (MI), Italy (E. Paulesu).
E-mail addresses: manuela.berlingeri@uniurb.it (M. Berlingeri), eraldo.paulesu@unimib.it (E. Paulesu).
Abbreviations: DEAR, differential empathic activation for race; IAT, Implicit Association Test; TPJ, temporo-parietal junction.

http://dx.doi.org/10.1016/j.neuroscience.2016.06.048
0306-4522/

INTRODUCTION

Empathy is the ability to understand and share other people's feelings; it has recently been defined as "feeling what another feels" (Gonzalez-Liencres et al., 2013). It is now maintained that empathy can be fractionated in at least two components (Rogers et al., 2007; Shamay-Tsoory et al., 2009), namely, "cognitive empathy" and "affective empathy". The former can be defined as the process of understanding another person's perspective, and as a consequence, it implies a certain level of awareness. Cognitive empathy has been recently included as part of the mentalizing process in the triangular model proposed by Zaki and Ochsner (2012). The latter, affective empathy, corresponds to the ability to emotionally respond to the affective state of others; this latter empathic response is more automatic and thus it does not necessarily require full awareness.

Of the different emotional states that may induce empathic reactions, those associated with pain have become a much-explored case study in cognitive neuroscience. Several physiological and neurofunctional studies recently showed that the empathic responses toward a third person in pain are associated with the activation of the same brain structures involved in first-person pain perception. These include the brain regions associated with motivational-affective dimensions of pain such as the anterior insular cortex and the anterior and mid-cingulate cortex (Singer et al., 2004; Singer and Frith, 2005); more recently, "empathic" responses have also been found in the nodes of the brain pain matrix (Peyron et al., 2000) that are closer to the incoming peripheral stimuli, such as the thalamic nuclei or the primary (SI) and secondary (SII) somatosensory cortices (Bufalari et al., 2007; Akitsuki and Decety, 2009; Aziz-Zadeh et al., 2012).

Interestingly, empathy, at least in some rudimentary form, is not unique to humans. In a recent review Gonzalez-Liencres et al. (2013) suggest that affective empathy may be present in animals, such as non-human primates that possess some form of self-awareness. From an evolutionary point of view, it has been proposed that empathy may be based on the ability that onlookers have to recognize an individual similar to them-selves while maintaining an alert status when in front of an individual belonging to a different species. Likewise, the empathic response in human beings can be moderated by contextual interpersonal variables (Cialdini et al., 1997; Baron-Cohen and Wheelwright, 2004; Hein and Singer, 2008; Hein et al., 2010) such as

the in-group/out-group social categorizations based on race differences between the target and the onlooker (see Cikara and Van Bavel, 2014 for a recent review). Evidence in favor of these findings (see Meconi et al., 2015 for a recent review on this topic) comes from electrophysiological studies (Forgiarini et al., 2011), Evoked Response Potential studies (Sessa et al., 2014a), TMS studies (Avenanti et al., 2010) and fMRI paradigms (Cunningham et al., 2004; Xu et al., 2009; Azevedo et al., 2012). For example, Cunningham et al. (2004) found a stronger activation for African over Caucasian faces in the amygdala, and this effect was directly proportional to the level of implicit racial bias as measured by the Implicit Association Test (IAT). The existence of racially biased empathic responses has also been confirmed by two recent fMRI studies (Xu et al., 2009; Azevedo et al., 2012). When participants observed an actor of their same race (in-group) receiving a painful stimulus (versus a harmless stimulus), the motivational/affective components of their cerebral pain matrix (e.g., the cingulate cortex and the insula) showed a greater activation than that for actors of a different race (out-group). This effect has been interpreted as a neural signature of inner racial biases for the in-group, and here we called it the "*differential empathic activation for race* (DEAR) effect" to make clear that this effect emerged from a significant race (in-group vs. out-group) by stimuli (painful vs. harmless) interaction effect in both of the aforementioned fMRI studies.

Furthermore, in the study by Azevedo et al. (2012), in line with Cunnigham's findings (2004) in the context of face-perception, the in-group DEAR effect during the presentation of the painful stimulus correlated with implicit measures of racism (Implicit Association Test – IAT race). However, the explicit judgment of the African actors' experience was, on average, similar to the explicit level of pain attributed to Caucasian actors. This result suggests that people may manifest controlled egalitarian behaviors toward members of the out-group, notwithstanding opposite implicit uncontrolled and embodied physiological reactions (as clearly suggested in the review by Eres and Molenberghs, 2013). Individuals show this behavior in an attempt to follow their desire to appear free from a racial prejudice, and some may use politically correct language in this regard.

To conclude, the pattern of results reported in the cognitive neuroscience literature is compatible with the idea that empathy develops from a mere embodied automatic, bottom-up process (shared with other species) to a more complex process also involving top-down modulatory control (as preliminarily suggested by the "temporary executive dysfunction in racially biased individuals", Richeson et al., 2003; Richeson and Shelton, 2003). As a consequence a full description of the empathic responses should take into account both automatic low-level and embodied responses, both higher level and cognitively mediated reactions, as suggested by Zaki and Ochsner (2012). This suggestion has been recently picked up by Sessa et al. (2014b) in a ERP study in which perceptual (i.e., painful or neutral facial expressions) and contextual (i.e., painful or neutral related sentences) cues on others' mental states were orthogonally manipulated. The results showed a temporal double dissociation of neural responses to others' pain: perceptual cues modulated the early activity at 110–360 ms over fronto-central and centro-parietal regions, whereas painful contexts modulated the late activity in the same regions.

Accordingly, we assume that, when an empathic response toward an actor from a different race is in order, two contrasting or complementary forces might be active, that is, a more primitive (from the evolutionary point of view) component based on uncontrolled automatic embodied responses biased in favor of in-group members and a more culturally evolved and cognitively dependent component based on the internalization and explicit application of social rules (as suggested also by Richeson et al., 2003).

The first component has been associated with the brain DEAR effect in the pain-matrix measured when people stare at stimuli designed to elicit an empathic response modulated by in-group/out-group factors (e.g., Xu et al., 2009; Azevedo et al., 2012). The second component may depend on the activation of top-down processes and the ensuing neural structures (e.g., the prefrontal cortices; Richeson et al., 2003; Amodio, 2014) that are necessary to explicitly generate, in developed societies, what are perceived as politically correct responses, i.e., actions or behaviors that are calculated on purpose to not offend or disadvantage out-group members.

In the present study, we explicitly explored the functional anatomical bases for the aforementioned dissociation. We expected to replicate the in-group-specific DEAR effect and to document a specific brain activity that is associated with the explicit behavior whereby *politically correct responses* are given when assessing the pain felt by members of an out-group.

EXPERIMENTAL PROCEDURES

Participants

Twenty-five normal Caucasian participants, 12 males (mean age = 25.3 years, SD = 4.81) were recruited among undergraduate university students and young workers. All participants gave their written consent to participate in the fMRI study. The study received the approval of the Ethics Board of the University of Milano-Bicocca (May 12, 2014, protocol number 126), and the procedures that we followed were in accordance with the Declaration of Helsinki (BMJ, 1991; 302:1194).

Racist trait measures

Before fMRI scans, participants sat in a quiet room and completed the Internal Motivation to Respond Without Prejudice Scale (IMS; Plant and Devine, 1998), an explicit self-assessment of a personal racist trait that typically correlates with other explicit measures of racisms. Participants also completed a Race (Caucasian and African) Implicit Association Test (IAT) to assess the implicit race biases in favor of African people or Caucasian people

Please cite this article in press as: Berlingeri M et al. *Guess who's coming to dinner*: Brain signatures of racially biased and politically correct behaviors. Neuroscience (2016), http://dx.doi.org/10.1016/j.neuroscience.2016.06.048