UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Psicologia dello Sviluppo e della Socializzazione

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE

CICLO XXVI

Dealing with others' physical pain reveals variance in empathic processes:

Evidence from event-related potentials.

Direttore della Scuola: Ch.mo Prof.ssa Francesca Peressotti

Supervisore: Dott.ssa Paola Sessa

Dottoranda : Federica Meconi

A tutte le mie persone empatiche

e all’empatia come prima delle qualità umane.

INDEX

RIASSUNTO

Nel presente elaborato sono descritti una serie di 5 esperimenti, condotti con la tecnica dei potenziali evento-relati (Event-Related Potentials, ERPs). E’ stata indagata la natura multisfaccettata dell’empatia umana per il dolore (Esperimento 1) e la sua variabilità in funzione della razza (Esperimento 2) e dell’affidabilità di un volto (Esperimenti 4 e 5). Il paradigma classicamente impiegato in questo tipo di studi è il compito di decisione del dolore in cui viene richiesto ai partecipanti di dire se l’individuo che si osserva (o parti del suo corpo) sta subendo una stimolazione dolorosa oppure neutra.

Inoltre, prima di indagare se l’affidabilità potesse modulare la risposta empatica, è stato condotto l’Esperimento 3 per testare se le caratteristiche fisiche di affidabilità fossero automaticamente estratte da un volto influenzandone l’elaborazione in memoria di lavoro visiva (MLV). E’ stato impiegato un diverso paradigma, il compito di rilevamento del cambiamento, e un correlato neurale diretto della risoluzione delle rappresentazioni in MLV.

La ricerca neuroscientifica sull’empatia al dolore si è focalizzata principalmente sulla natura multicomponenziale di questa capacità. Sono stati identificati almeno due sotto processi dell’empatia anatomicamente dissociati: l’*experience sharing* e il *mentalizing*. Il primo include gli aspetti più affettivi e legati al contagio sensorimotorio che permettono di sentire internamente lo stato emotivo dell’altro; il secondo si riferisce alla capacità di inferire gli stati mentali altrui. Un importante scopo del presente elaborato è quello di fornire evidenza empirica di una loro possibile dissociazione anche funzionale, nel dominio temporale.

Nell’Esperimento 1 è stato indagato questo aspetto implementando una specifica versione del compito di decisione del dolore. Ai partecipanti sono state presentate sia

un’informazione sensorimotoria (una faccia con espressione neutra o di dolore) che un’informazione contestuale (la descrizione di un contesto neutro o di dolore) nella stessa prova sperimentale per evidenziare il decorso temporale della reazione elettrofisiologica al dolore legata ad entrambi i sotto processi dell’empatia. Si è così dimostrata una selettiva attivazione dell’experience sharing e del mentalizing in due diverse finestre temporali.

Sulla base di questi risultati è stato possibile esplorare la variabilità della risposta empatica a diversi tipi di stimoli esterni. Precedenti studi hanno suggerito che le persone sono più empatiche nei confronti del dolore subito da un individuo appartenente alla nostra stessa razza rispetto ad individui appartenenti ad altre razze (Avenanti et al., 2010; Xu et al., 2009). Nell’Esperimento 2 è stato dimostrato che questa risposta preferenziale è confinata ai meccanismi dell’experience sharing. Il mentalizing mostra infatti una risposta empatica anche verso il dolore di individui di un’altra razza.

Sebbene la valutazione della razza di un volto sia implicita, questa viene elaborata in modo automatico e veloce sulla base delle caratteristiche fisiche di un volto. Recentemente è stato dimostrato che anche la valutazione di affidabilità di un volto avviene a prima vista (Willis e Todorov, 2006) similmente alla razza. E’ stato così ipotizzato che l’affidabilità, sia in volti computerizzati (Esperimento 4) che in volti reali (Esperimento 5) modulasse l’empatia, anche in assenza di una conoscenza della personalità di dell’individuo o del suo comportamento sociale, perché questa può implicitamente e velocemente influenzare le nostre interazioni sociali. Per determinare l’efficacia della valutazione di affidabilità (Oosterhof and Todorov, 2008), nell’Esperimento 3 è stato testato l’impatto delle caratteristiche fisiche standardizzate di affidabilità sulle rappresentazioni in MLV, anche quando irrilevanti ai fini del compito.

SUMMARY

The present work consists of a review of 5 event-related potentials (i.e., ERPs) experiments I conducted, which deal with the multifaceted nature of human empathy for pain (Experiment 1) and variances in empathic processes, as a function of others' race (Experiment 2) and others' perceived trustworthiness, i.e. driven by facial features (Experiments 4-5), addressed through classical and modified versions of the pain decision task. The classical version of the pain decision task requires participants to decide whether presented stimuli (either pictures of individuals or body parts) receive either painful or neutral stimulation.

Furthermore, prior to investigate trustworthiness as modulator of neural empathic response, I adopted in Experiment 3 a different paradigm, namely the change detection task, and a direct neural correlate of the resolution of visual working memory (i.e., VWM) representations to test whether trustworthiness is automatically extracted from faces biasing VWM processing.

The main issue of the neuroscientific research on empathy for pain is about its multiple aspects. Indeed, neuroscientific research identified at least two subprocesses constituting empathy: Experience sharing and mentalizing. The former encompasses affective and sensorimotor aspects to inner feel the other's emotive state; the latter allows to infer/attribute the other's mental state. Experience sharing and mentalizing appear to be at least anatomically dissociated. One important aim of the present thesis is to provide evidence on the possible functional dissociation in the temporal domain.

In Experiment 1 I addressed this issue by implementing a new version of the pain decision task. I presented participants with both sensorimotor (picture of a face with either painful or neutral expression) and contextual information (a sentence describing either a painful or neutral context) to highlight the deployment of electrophysiological

reaction to pain related to the both subprocesses and I provided evidence of selective engagement of experience sharing and mentalizing into two time-windows.

This is the starting point of the present studies on the way of exploring variance in neural empathic response. Previous studies suggested that people are more naturally empathic towards own-race individuals relative to other-race individuals (Avenanti et al., 2010; Xu et al., 2009). In Experiment 2 I provided compelling evidence that such preference is confined to experience sharing. Indeed, mentalizing is responsive to other-race pain.

Although implicitly appraised, race of a face is processed quickly and automatically driven by physical facial features. Recently it has been demonstrated that evaluation of perceived individuals' facial trustworthiness is appraised at first sight (Willis and Todorov, 2006), similarly to race. I hypothesized that trustworthiness, either in computerized faces (Experiment 4) and real faces (Experiment 5) plays another key role in modulating empathy even in the absence of previous knowledge on others' personality and social behavior because it can implicitly and quickly shape our social interactions. In an attempt to determine the efficacy of trustworthiness appraisal, I tested in Experiment 3 whether and how standardized physical facial features of trustworthiness (Oosterhof and Todorov, 2008) bias VWM processing even when task-irrelevant.

Chapter 1

Neuroscience of empathy

Empathy in a shot

Over millenniums, the evolution of our brain arose in a context of continuous interactions among individuals organized in social groups, it follows that it is doubtless specialized in such capacity. Empathy allows people to be connected and react appropriately in social interactions by relating to and understanding the emotional states of the others, likely but not necessarily, either for prosocial behaviour (i.e., the motivation to improve others' experiences by, for instance, reducing others' sufferance, e.g. de Waal, 2008) or reducing personal distress. Whether it is permanently true is the broad question of the current studies.

Social neuroscience laboratories deeply investigated empathy, predominantly with fMRI, in several contexts such as, for instance, empathy to disgust (e.g., Benuzzi, Lui, Duzzi, Nichelli, & Porro, 2008; Wicker Keysers, Royet, Gallese & Rizzolatti, 2003) to fear (de Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004), to neutral touch (e.g., Keysers, Wicker, Gazzola, Anton, Fogassi & Gallese, 2004) and to others' pain, which is the object of the current studies.

Social neuroscientists specifically tested empathy to others' pain by manipulating either the affective relationship between the perceiver and the target, i.e. lovers (Singer, Seymour, O'Doherty, Kaube, Dolan & Frith, 2004), or the social fairness of the target (Singer, Seymour, O'Doherty, Stephan, Dolan & Frith, 2006); by manipulating either the contextual reality of stimuli (Gu & Han, 2007) or facial expressions (Botvinick, Jha, Bylsma, Fabian, Solomon & Prkachin, 2005). More recently, also the racial group of both the perceiver/observer and the target was investigated as a potential modulator of empathy (e.g., Avenanti, Sirigu, & Aglioti, 2010; Chiao & Mathur, 2010; Forgiarini, Gallucci, & Maravita, 2011; Sheng & Han, 2012; Xu, Zuo, Wang, & Han, 2009)

But what is empathy? People commonly define empathy as that complex isomorphic capacity to feel, vicariously, the other's emotional state *with* the other. Indeed, empathy can be considered that capacity that makes the interaction between people possible by reading others' facial expressions and understanding intentions or, more subtly, *potential* intentions, which appear to be an additional step for a full-blown empathic experience.

And what is not empathy? Empathy is not merely emotional contagion, which is the tendency to adopt others' emotional states; it is not simple mimicry, which is the automatic synchronization of our affective expressions, gestures or vocalization with those of an observed individual; empathy can be followed by sympathy, compassion and empathic concern that, as opposed to empathy per se, are other-oriented (for a more detailed distinction between these terms, please see Singer & Lamm, 2009).

In the next sections, I'll go into deeper detail about a neuroscientific definition of empathy. I'll first analytically discuss existing literature on empathy and on empathy towards others' pain, which is the object of the present thesis. Secondly, I'll go into more detail of the neuroimaging technique I adopted to conduct the current studies and I'll describe evidence on empathy for pain, which used the same technique: Event-

Related Potentials (i.e., ERPs). Thirdly, I'll critically discuss the variability of the empathic process and, fourthly, the promises and the pitfalls of this body of research. Lastly, I'll provide an overview of the current studies.

1.1 Neuroscience of empathy

Several theoretical models have been proposed in social and cognitive neuroscience domains in an attempt to define the nature of empathy and its underlying mechanisms.

Mirror neurons. Some scholars proposed that empathy starts with motor resonance by means of mirror neuron system (Gallese, 2001), which allows internal simulation for action understanding. Mirror neurons are a class of neuronal cells first discovered in the rostral area of macaques' ventral premotor cortex (F5 area) by di Pellegrino and colleagues in 1992 (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) and then termed *mirror neurons* in a subsequent study by Gallese and colleagues in 1996 (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). In that early study, single-cells activity was recorded from electrodes positioned inside the monkeys' brain while observing an experimenter grasping some food. Surprisingly enough, these cells seem to have the property to discharge when a specific goal-directed motor act is either executed or observed by the monkeys. Indeed, the mere observation of either an object, a mimed or a non-goal-directed (e.g. intransitive) action was not sufficient to trigger mirror response. Importantly, one may argue, against the early evidence of the existence of mirroring, that monkeys exposed to some food might have been preparing the motor act of grasping for hunger, and reported neurons discharge was related to that instead of being related to the observation of such action by an experimenter. This argument can be ruled out by the evidence that the food per se does not trigger mirror neurons discharge. So does not another interesting object. Crucially, the fundamental aspect for mirroring is the interaction between the agent and the target by means of a biological effector (either the mouth or a hand). Thus, it's the observation of a realistic goal-directed (i.e. transitive) action toward the food that triggers mirror response in monkeys. Such evidence suggests that mirror neurons are involved in action understanding. In this regard, Umiltà and collaborators (Umiltà, Kohler, Gallese, Fogassi, Fadiga, Keysers, & Rizzolatti, 2001) demonstrated that mirror neurons of F5 area discharge even when monkeys can't see the occurring action although they can infer it (Figure 1).

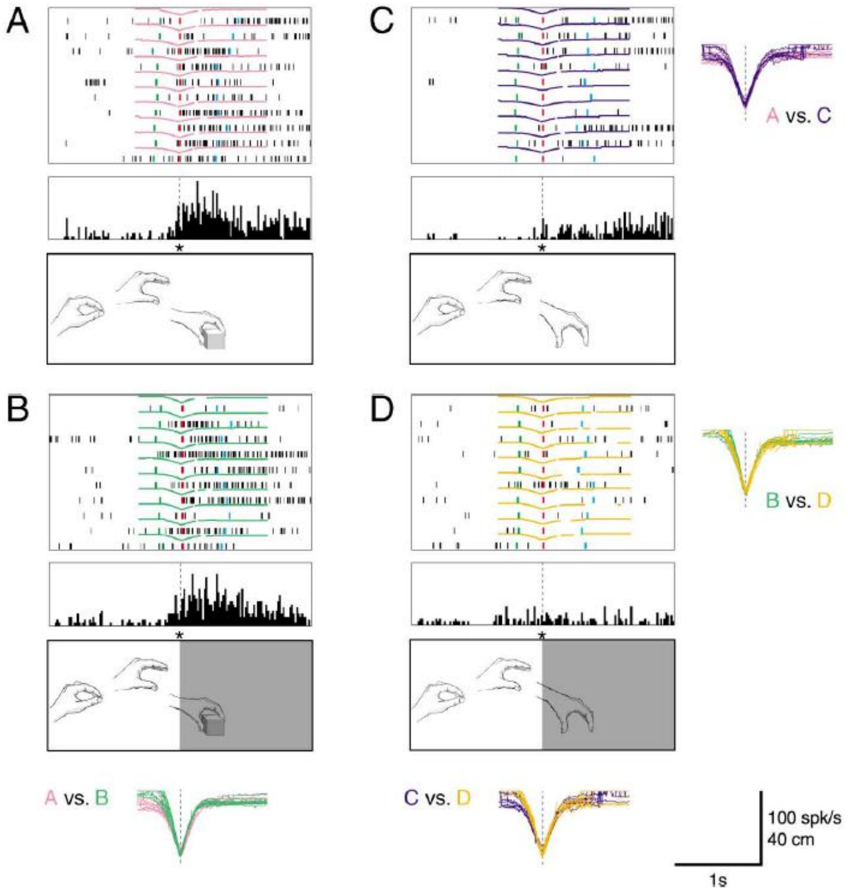


Figure 1. Adapted from Umiltà et al., 2001. Example of a neuron responding in action observation in full vision (A) and hidden condition (B) but not in the mimed conditions (C and D).

Over these last twenty years, single-cell recordings studies have reported mirror neurons not only in the rostral division of the ventral premotor cortex (i.e., vPMC) in macaques but also in the dorsal premotor (i.e., dPMC) and primary motor cortex (i.e., MC) and in the inferior parietal lobule (i.e., IPL; for reviews, please see e.g., Rizzolatti

& Craighero, 2004; Rizzolatti, Fabbri-destro, & Cattaneo, 2009).

Although there is no single-cell recordings evidence of mirror neurons in humans, there is a huge amount of studies suggesting the existence of such neuronal cells system in humans. Neuroimaging data demonstrated that the circuit of mirror neurons in humans involves the inferior parietal lobule (IPL), the lower part of the precentral gyrus (i.e., PCG) and the posterior part of the inferior frontal gyrus (IFG) including the Broadmann Area (BA) 44, i.e. the *pars opercularis* (i.e., PO), which is the homologue of macaques' F5.

Notably, in adult humans the homologue mirror neuron system is active also for intransitive movements, i.e. those movements that are not goal directed or are meaningless. Fadiga et al. (1995) used transcranial magnetic stimulation[1] (i.e., TMS) to stimulate the motor cortex while motor evoked potentials[2] (i.e., MEPs) were recorded during the observation of either a transitive, an intransitive action and a control condition. The observation of both transitive and intransitive action, relative to the control condition, elicited, in the observer, a selective increase in MEPs recorded from the corresponding muscle being observed, supporting then sensorimotor contagion in the observer as index of empathic reaction to others' pain.

Electrophysiological studies showed evidence of mirror activity in the desynchronization of mu rhythm, which is a particular rhythm of the alpha band (8-15 Hz) registered from the motor cortex with the electroencephalography (i.e., EEG) over the central sites. Mu rhythm desynchronization manifests when an individual observes an action done by another individual (e.g., Cochin, Barthelemy, Roux, & Martineau, 1999; Gutsell & Inzlicht, 2010; Oberman, Hubbard, McCleery, Altschuler, Ramachandran, & Pineda, 2005); a similar result has been observed also in infants

---

[1] In brief, the TMS is a non-invasive method that uses focal electromagnetic induction to induce weak electric currents with rapidly changing magnetic field. This induction causes the alternative depolarization and hyperpolarization of the neuron cells under the coil used to induce electric currents.

[2] The registration of electrical activity through the placement of specific electrodes along the muscles of interest.

(Nyström, Ljunghammar, Rosander, & von Hofsten, 2011; Nyström, 2008). Notably, similarly to monkeys, only transitive, and not intransitive, movements trigger mirror activity in infants, this can be taken as suggestion of primitive stages of mirror system that finally develops in a more sophisticated form of action understanding and other forms of intersubjective sensitivity, such as empathy, for instance. However, other lines of reasoning suggest that an alternative hypothesis is that action understanding of intransitive movements might be related to the activation of mirror neurons mediated by mentalizing, i.e. mindreading, which is a more cognitive subprocess of empathy that I'll discuss in the following sections (Goldman & Jordan, 2013).

Gallese (2003) proposed that mirror-like neural response is based on shared brain areas for first- and third-hand experience. Neuroimaging studies on humans showed that mirror mechanism is not confined to the premotor system but it is present also in those cortical circuitries that mediate visceromotor and emotion-related behaviors such as observing or firsthand feelings. Wicker and co-workers (Wicker et al., 2003) showed, for instance, that the anterior part of the insula (i.e., AI) is involved in both feeling and seeing disgust.

The insular cortex has an integrative role in linking information from several functional systems and is characterized by anatomical distinctions, a recent meta-analysis revealed that it shows four functionally distinct regions: the anterior-dorsal region is involved in processing some cognitive functions; the right middle insular gyrus is specific for the olfacto-gustatory-domain; the mid-posterior insula can be activated by tasks from the sensorimotor domain; finally the anterior-ventral insula is associated with socio-emotional functions, such as emotional processing and empathy (Kurth, Zilles, Fox, Laird, & Eickhoff, 2010). The insular cortices are a center of integration of interoceptive information; in the case of empathy, it can perceive and internally represent external stimuli; according to several psychophysiological models the experience of emotions depends on changes in bodily arousal states (e.g., Craig, 2003; 2009). Specifically, it has been suggested that the anterior-ventral insula is the link between mirror neuron system and emotional processing, allowing empathy to others' emotions through underlying mechanisms of emotion understanding (Iacoboni & Dapretto, 2006; Iacoboni, 2009a; Singer, Critchley, & Preuschoff, 2009).

Singer and collegues (2004; 2006) showed that the bilateral AI together with the anterior cingulate cortex (ACC) are involved in both experiencing and observing physical pain.

Whereas the insula is a center of representation and integration of others' internal and emotional feelings, the ACC has been proposed to be the motivational and action-related counterpart (Craig, 2002; 2009). Of course, the ACC is not an "empathy region" per se but it is related to the abstract and subjective feeling of unpleasantness, or more specifically the unpleasantness of pain (Amodio & Frith, 2006; Kong, Gollub, Polich, Kirsch, LaViolette, Vangel, Rosen, & Kaptchuk, 2008). It is part of the limbic system and it acts as a hub for affective, cognitive and motor control functions (Paus, 2001). Several studies proposed the involvement of the ACC in pain processing (Craig, 2003a;b; Dum, Levinthal, & Strick, 2009; Qiu, 2005); Hutchinson and colleagues (Hutchison, Davis, Lozano, Tasker, & Dostrovsky, 1999) corroborated such view by identifying single neurons in the ACC that respond selectively for painful stimuli, in this vein, Iacoboni (2009) suggested that these neurons might act as pain-specific mirror neurons. The International Association for the Study of Pain (i.e. IASP) defined pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (IASP, 1994). First-hand experience of pain is endowed with three dimensions: A sensory-discriminative, related to bodily location, intensity and duration of pain; an affective-motivational level, related to the unpleasantness of nociceptive stimulus; and a cognitive level, related to the cognitive evaluation of pain and the class of motor planning aiming at reducing pain. Indeed, the functional magnetic resonance imaging[3] (i.e., fMRI) study conducted by Singer and co-workers (e.g., Singer et al., 2004) provided evidence in line with this. They tested empathy towards others' pain in lovers: Female partners were lying in the scanner next to their partner, which were taking part of the study by seating in the same room. Women either received a painful or non-painful stimulation or observed a light signaling that their partner would receive the same stimulation, either painful or non-painful. The experience of perceiving pain in first person activated the pain-related network known as "pain-matrix" (e.g., Singer et al., 2004). This network consists of brain areas coding for sensory-discriminative, affective-motivational and evaluative cognitive dimensions of pain experience described above (Figure 2).

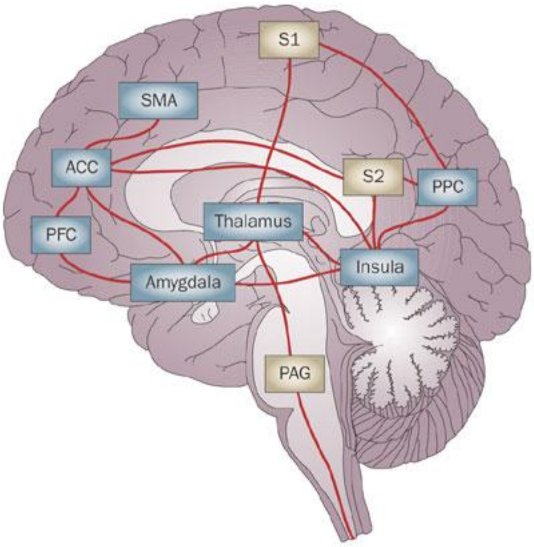


Figure 2. This so-called pain matrix mainly involves the thalamus, the amygdala, the insula cortex, the supplementary motor area, the posterior parietal cortex, the prefrontal cortex, the cingulate cortex, the periaqueductal gray, the basal ganglia and cerebellar cortex (not shown), and the primary and secondary sensory cortex. Abbreviations: ACC, cingulate cortex; Insula, insula cortex; PAG, periaqueductal gray; PFC, prefrontal cortex; PPC, posterior parietal cortex; S1, primary sensory cortex; S2, secondary sensory cortex; SMA, supplementary motor area. Taken from Springer Science + Business Media: Neurological Sciences, Neuroimaging:

---

[3] In brief, the functional magnetic resonance imaging (i.e., fMRI) is a procedure that uses magnetic resonance imaging technology to measure brain activity by detecting associated changes in blood-oxygen-level-dependent (i.e., BOLD) signal while participants are performing a cognitive task. Its spatial resolution is about 3 mm^3 and each scan can be acquired in less than 1 minute.

visualizing the brain in pain, volume 28, 2007, May.

Notably, in Singer et al.'s study, the first-hand experience of pain activated the somatosensory node of the pain matrix, which, again, processes the bodily location and intensity of pain, whereas third-hand experience did not. Crucially, in both trials when women were receiving painful stimulation or were signaled their partners were receiving painful stimulation, the affective-motivational node of the pain-matrix was activated. Specifically, the anterior medial cingulate cortex (ACC), the anterior insula, and the cerebellum. Interestingly, these regions are involved in processing the desires, urges, or impulses to avoid the painful experience and shared the first- and third-hand experience of pain (Singer et al., 2004; 2006). However, the absence of the sensorimotor node of the pain matrix is in contrast with TMS evidence as it has been shown by Avenanti and collegues (Avenanti, Bueti, Galati, & Aglioti, 2005). I will describe this contradictory evidence below.

Gallese's *shared manifold hypothesis* has been suggested to be at the basis of the intersubjectivity that conveys empathy. Consistently with the Perception-Action coupling Model (PAM)[4] theorized by Preston and de Waal (2002), witnessing the emotional state of someone determines the inner representation of such state automatically and unconsciously. The closer are the observer and the perceiver, the stronger would be such coupling (e.g., de Waal, 2008; Singer et al., 2004; Xu et al., 2009). The modulation of the strength between perceptual input and the corresponding representation is also a function of individual differences in empathy capacity. An intriguing evidence of this interplay was provided by an fMRI study of Gazzola, Aziz-Zadeh and Keysers (2006). Some participants were selected on the basis of their scores

---

[4] The Perception-Action coupling Model posits that seeing an object's state lead to automatically form in the observer a corresponding representation, which in turn, finally activates somatic and autonomic responses.

obtained in an empathy questionnaire, i.e. the Interpersonal Reactivity Index (i.e. IRI; Davis, 1980, 1983). This self-report questionnaire (see Appendix 1) is composed of four 7-items subscales that measure both the capacity of taking others' perspective (Perspective Taking, i.e. PT, and Fantasy, i.e. F, subscales) and the capacity to feel with the others their feelings (Empathic Concern, i.e. EC, and Personal Distress, i.e. PD, subscales). Responses are given of 5-point scale ranging from 1 to 5. The authors observed that participants with high scores of Perspective Taking showed stronger mirror activation whereas participants with low scores on the same subscale showed no significant mirror activity suggesting that mirroring seems to interact with mechanisms involved in goals understanding (also defined as *mentalizing)* and others' motivation rather than emotional sharing. However, a recent meta-analysis showed that the mirror neuron system and the mentalizing are rarely concurrently active, so they do not interact but they are rather complementary (Van Overwalle & Baetens, 2009).

Such result opens a crucial issue on the way of defining empathy: the multiple aspects of such cognitive process, namely the capacity of either internally simulating and mentalizing others' emotional state and whether and how they are related for a full-blown empathic experience.

Mentalizing. The *mentalizing* refers to that process that allows to cognitively infer the others' mental state through diverse modalities, e.g. by either observing the other's facial expression, through the contextual appraisal or our knowledge about the other. It is a subprocess of empathy. Most of the neuroimaging studies that investigated the cognitive aspects of empathy asked participants to imagine what the observed target person was either thinking or feeling, other studies investigated how individuals are able to take the other's perspective.

An exemplar study in this particular domain is the one of Ruby and Decety (2004), in which they explored a particular aspect of mentalizing, namely the ability of taking the other perspective.

That was the third of two prior positron emission tomography[5] (i.e., PET) studies (Ruby & Decety, 2001; 2003) in which the authors instructed participants to adopt their own perspective or the perspective of another person in performing an action (i.e., motor domain) and in responding to a medical question (i.e., conceptual domains, this study was conducted with medical students). The main claim of these studies was that individuals need to distinguish the self from the other in order to activate the mentalizing system and to correctly attribute the agency to the other. Indeed, their findings revealed the involvement of brain areas largely involved in the self-representation, such as the IPL and the precuneus other than the posterior cingulate and the frontopolar cortex. Ruby and Decety (2004) conducted this third study in order to extend their results to the emotional domain. The authors asked participants, selected as medical students, to adopt their own perspective or the perspective of their mother in neutral or emotional condition. In the neutral condition, they were requested to give an opinion on the reliability of some declarations in the medical domain; in the emotional condition, participants were requested to report the emotional reaction that some real-life situations would induce. Results replicated previous ones and extended them to the emotional context: the first person relative to the third person perspective showed the activation of the somatosensory cortex, whereas the right IPL and the ventromedial prefrontal cortex (i.e., vmPFC) including the frontopolar cortex and the gyrus rectus were activated in the reverse comparison, regardless of the emotional content of situations presented. Brain areas involved in emotional processing, including the amygdala and the temporal poles, were activated in the emotional relative to the neutral

---

[5] The positron emission tomography is a nuclear medicine, functional magnetic technique. The system detects pairs of gamma rays emitted indirectly by a positron-emitting-radionuclide (i.e., tracer) which is introduced into the body on a biologically active molecule. Through the detectors is possible to calculate activation of neural populations in different brain areas while participants are performing a cognitive task. 3D images of tracer concentration within the body are then constructed by computer analysis through subtractive method. Its spatial resolution is 5-10 mm^3, each scan needs several minutes to be acquired.

condition, regardless of the perspective participants adopted.

As I mentioned above, the crucial aspect of mentalizing is the representation of the mental states of the others, by attributing/inferring them. Commonly, participants perform “false beliefs” tasks to test whether individuals can form a belief about the other in discrepancy with what they know to be the truth. In these kind of tasks, participants read some stories or rather are exposed to a figurative representations of them and are required to say what the characters of the stories think. Indeed, in these stories the main character puts something in a box, then leaves the room and in its absence another character takes it and leaves the room. Finally, the main character returns in the room. Participants’ task is to say what is the belief of the main character, that is, what the main character will do or think (e.g., “Will the main character look in the box to take what put inside?”).

Several reviews and meta-analysis, vastly reported that this kind of tasks involve frontal circuitries (Frith & Frith, 2003; 2006; Frith & Singer, 2008; Van Overwalle & Baetens, 2009; Van Overwalle, 2009). Frith and Frith (2006) reviewed that other brain areas involved in mentalizing are a sub-set of temporal regions such as the posterior end of the superior temporal sulcus (pSTS), and the adjacent temporo-parietal junction (TPJ); the former is involved in gaze-cueing (Pelphrey, Morris, & McCarthy, 2004) and in visual detection of the biological motion (Gao, Scholl, & McCarthy, 2012), and the latter is involved in the representation of the body in the space (Blanke, Mohr, Michel, Pascual-Leone, Brugger, Seeck, Landis, & Thut, 2005) and in the identification of human goals and intentions (Van Overwalle, 2009). These functions make observers direct their attention to the context and to change perspective by distinguishing the self from the other. Other regions involved in mentalizing are the precuneus, which is also implied in self/other distinction (e.g., Ruby & Decety, 2001) and the temporal pole (i.e., TP), which is activated during memory retrieval from autobiographical memory

(Maguire, Mummery, & Büchel, 2000); Frith and Frith (2003) suggested that this last region might help in generating a semantic and emotional context that allows the interpretation of stories and pictures that are currently being processed, on the basis of personal past experience.

In the light of these early studies, several neuroscientists agreed that empathy is the resulting process of bottom-up, stimulus driven and automatic, and top-down, controlled, mechanisms (de Vignemont & Singer, 2006; Decety & Lamm, 2006; Hein & Singer, 2008; Singer & Lamm, 2009; Singer, 2006).

Decety and Jackson (2006) proposed a first model in which three major components of empathy are identified despite several definitions:

- *Affective response to other's emotional state*. Consistently with Singer's lab findings (2004; 2006) this component endorses only an affective nature and does not count a sensorimotor contagion. Importantly, it is thought to be automatically activated by perceptual input that directly matches with action in a bottom-up process. This view is related to the mirror neuron system theory, the perception-action coupling model by Preston and de Waal (2002) and shared neural bases for first and third-hand emotional experience hypothesis;
- *the cognitive capacity to represent the other's emotional state* by taking its perspective and inferring the other's mental state. It entails the involvement of executive resources that top-down regulate empathy;
- *emotion regulation*, which implies the capacity to disentangle the observer from the other in order to prevent a total overlapping between representations of the self and the other and so provoking distress or anxiety.

Sensorimotor contagion. Recently, Avenanti and collegues (Avenanti et al., 2005) revealed with a TMS study that the sensorimotor cortex is active during observation of pain in others. They registered MEPs induced by TMS stimulation of the left motor cortex while healthy participants observed a needle penetrating either the hand or the feet of a model. The authors found a reduction of corticospinal excitability that was specific to the muscle that participants observed being penetrated (for similar results, please see also Avenanti, Minio-Paluello, Sforza, & Aglioti, 2009). This findings seem in contrast with fMRI studies of empathy toward others' pain but a recent meta-analysis showed that such discrepancy might be due to differences in experimental paradigms (Lamm, Decety, & Singer, 2011). Indeed, in those fMRI studies, which used a paradigm similar to that used by Avenanti and colleagues (2005), sensorimotor activation is observed. Crucially, the class of neuroimaging studies on empathy investigated the different aspects of empathy by presenting participants either with coloured arrows signalling that another person was receiving a painful stimulation (i.e., cue-based paradigms) or with pictures representing a painful or non-painful stimulation (i.e., picture based paradigms). The cue-based paradigms require participants to imagine (i.e., to mentally represent) the painful stimulation inflicted to the other after being signalled about it. That might reliably be more related to the mentalizing dimension of empathy. Picture-based paradigms, such as the one used in Avenanti and collegues' studies (2005; 2009), requires the direct observation of the painful stimulation and it might be more related with the sense of sharing, which may involve the sensory contagion.

Recent model of empathy. An important consequence of this interesting evidence was that empathy has been enclosed in a multicomponential model endowed with an affective/sensorimotor and a cognitive nature. Zaki and Ochsner (2012) proposed a new model of empathy that mainly refers to:

- *Experience sharing*: The abilities of sharing others' experience as a neural resonance mechanism (i.e., affective/sensorimotor component)
- *Mentalizing*: The ability of representing other people's internal states by taking others' perspective (i.e., cognitive component).

Although several promising findings, it is still not clear whether or how experience sharing and mentalizing are interrelated or rather completely dissociated. Neuroimaging studies evidence an anatomical dissociation: The experience sharing has neural underpinnings in the mirror neuron system (i.e., IPL, IFG, and dPMC), and in the limbic system (i.e., amygdala, ACC, AI and ventral striatum; e.g., Keysers, Kaas, & Gazzola, 2010; Lamm & Singer, 2010; Rizzolatti & Sinigaglia, 2010). The mentalizing has been shown to be originated in prefrontal cortical circuitries (Decety & Jackson, 2006; Decety & Lamm, 2006), with neural underpinnings in the dorsomedial, dorsolateral, and medial prefrontal cortices (DMPFC, DLPF, and MPFC, respectively), middle frontal-gyrus (MFG), temporo-parietal junction (TPJ) and precuneus (e.g., Amodio & Frith, 2006; Decety, 2011; Lamm et al., 2011; Saxe & Kanwisher, 2003). Figure 3 depicts a schematic representation of the anatomical dissociation between experience sharing and mentalizing.

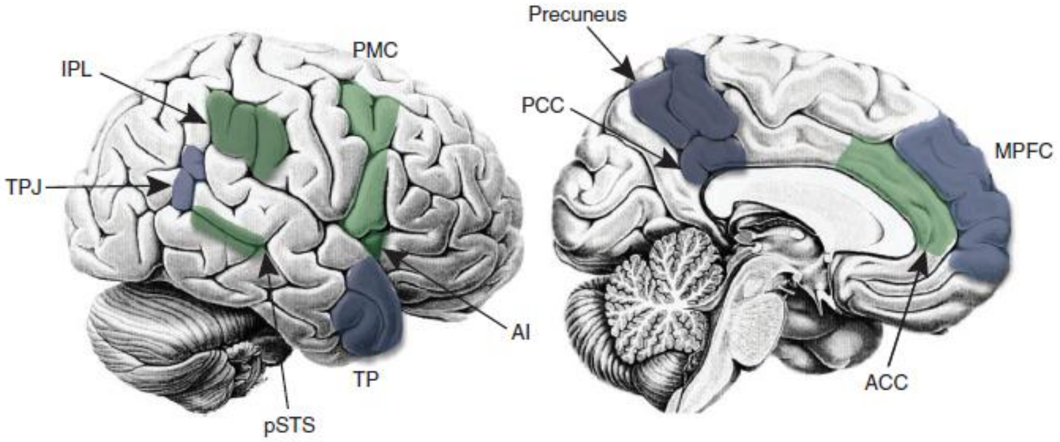


Figure 3. Schematic representation of the anatomical dissociation between experience sharing (green areas) and mentalizing (blue areas).

A precious empirical landmark of such anatomical dissociation came from a lesional study by Shamay-Tsoory and collegues (Shamay-Tsoory, Aharon-Peretz, & Perry, 2009). The authors individuated patients with lesions in either the IFG (i.e. a core region of the experience sharing) or in the vmPFC (i.e. a core region of the mentalizing system) and administered them with two empathy tasks, one for measuring experience sharing and the other for measuring the capacity of mentalizing. Their results revealed that IFG patients showed a selective deficit in experience sharing whereas vmPFC patients showed a selective deficit in mentalizing.

Perspective. Although research on empathy demonstrated that human brain can understand and empathically react to others' facial emotions such as disgust (Gallese, et al., 2004) fear (de Gelder et al., 2004), happiness (Foroni & Semin, 2011) or emotional suffering (Cheon, Im, Harada, Kim, Mathur, Scimeca, Parrish, Park, & Chiao, 2011), empathy towards others' physical pain research arose fundamental findings in improving research on empathy. That's because of the objectivity of receiving painful stimulation, which allowed a clear distinction of the brain areas active in the first person experience of pain and those active in observing pain (e.g. Avenanti et al., 2005; Ruby & Decety, 2004; Singer et al., 2004).

Empathy to others' physical pain research suggested an anatomical dissociation between experience sharing and mentalizing but less is known about functional dissociation and the time-course of the neural empathic response to others' pain (e.g., Spunt & Lieberman, 2012). Indeed, fMRI and TMS techniques do not offer a good temporal resolution. In the current studies the Event-Related Potentials (i.e., ERPs) technique has been used by virtue of its excellent temporal resolution (i.e., 1 ms) in order to track the time-course of the cognitive processes involved in empathy and so unravelling possible functional dissociation between them.

In the light of the studies discussed in this section, empathy is an ability composed of at least two subprocesses: the experience sharing, which encompasses affective and sensorimotor aspects to inner feel the other's emotive state, and mentalizing, which allows to infer/attribute the other's mental state. Experience sharing and mentalizing appear to be at least anatomically dissociated. One important aim of the present thesis is to provide evidence on the possible functional dissociation in the temporal domain.

1.2 Electrophysiological studies of empathy to others' physical pain

ERPs are revealed through the registration of the electroencephalogram (i.e., EEG). The EEG is used to monitor and graphically observe oscillations of the electrical potential[6] originated from the spontaneous electrical activity of the cerebral cortex, which is captured on the scalp. The EEG is obtained by placing some electrodes on participants' scalp. It reveals the total amount of the synchronized excitation in the neural population underlying each electrode; in particular the sum of post-synaptic potentials of pyramidal cells.

ERPs are a technological evolution of the EEG and they are fluctuations of the electrical potential in response to a sensory stimulation, associated with a psychological process and in preparation of motor activity. Crucially, ERPs are strictly in temporal relation with the presented event. The registered waves contain the so-called components, which are non-aleatory oscillations and rather have specific waveforms, polarity (i.e. either positive or negative), latency and scalp distribution; noteworthy, they can be classified as either early, intermediate or late depending on their latency. Importantly, they reflect on a sort of virtual continuum the perceptual process of the sensory stimulation, i.e. the earliest components, to higher level cognitive process, such as attention, memory, the mid- and long-latency components. Figure 2 offers a synthetic

---

[6] Electrical Potential = the force exerted on a particle with electrical charge, it is represented with the V symbol and it is measured in volt, i.e. voltage or potential difference. It is important to pinpoint that an ERP waveform reflects the active and reference site, which can be placed in either an active electrode, too, in an inactive electrode, such as the earlobe, or computed as the average of all the active electrodes, i.e. the *average reference* (for more details, see for example Picton et al., 2000). In the studies described in the chapter, the average of the electrodes placed on the two earlobes have been used.

view of some of a waveform including some of the most common ERP components: The P1[7], N1, P2, N2 and the P3.

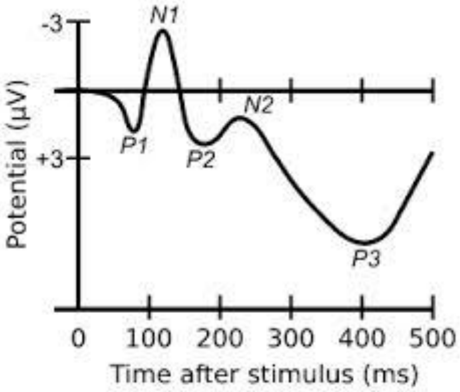


Figure 4. A synthetic view of a waveform including some of the most common ERP components: The P1, N1, P2, N2 and the P3. Note that the negative voltages are plotted upward, a common, but obsolete, practice in ERP research.

In the context of electrophysiological studies on empathy for pain some specific mid-latency (i.e., N1, P2, N2, and N3) and long-latency (i.e., P3) ERP components are involved.

The electrophysiological empathic response to others' pain manifests as a more positive deflection in response to painful relative to neutral stimulation registered in all the observed components, either positive or negative.

The N1, the P2 and the N2 are sensitive to attentional manipulations. In particular, the N1 and the P2 amplitudes are enhanced for unfamiliar stimuli, whereas the N2 amplitude is grater for familiar stimuli (e.g., Luck & Kappenman, 2011). For instance, when participants search for targets defined by specific features or combinations of

---

[7] The capital letter is related to the polarity of the components: "N" is for negative and "P" for positive. The number can be related to the serial number in which the oscillation occurs within the waveforms (e.g. the N1 is the first negative oscillation after the presentation of an event) or rather can be related to the latency, expressed in milliseconds, of the component (e.g. N100). The two way of classifications are not completely equivalent. Indeed, the first way of classification does not take into account the latency. Depending on the experimental design, the process the component is reflecting, and the sensorial modality in which the stimuli are presented, the deflection can occur with different latency range. The second way of classification pinpoints the temporal sequence in which the deflection occurs.

features in a visual search paradigm[8], a modulation of the P2 and the N2 amplitudes is observed.

Luck and Hillyard (1994a) required participants to perform a visual search task in which a pop-out feature was designated as target at the beginning of each block. Participants had to respond as fast as possible to the presence of the target pop-out resulting in three experimental conditions: The target pop-out, the nontarget pop-out (i.e., the pop-out feature was not the one designated at the beginning of the block) and the homogeneous condition (i.e., no pop-out features). At frontal sites, the N2 was larger for both the target and nontarget conditions when compared to homogeneous arrays. By contrasts, the P2 was larger only in response to target pop-out relative to nontarget and homogeneous conditions. So, whereas the N2 appears to be related to an automatic detection of popout stimuli, irrespective of their relevance for the task; what appear to be important in the P2 modulation is the presence of the relevant feature to be attended, irrespective of the presence or absence of irrelevant, although popout, features. Thus, the P2 reflects the detection of a specific feature.

Interestingly, the modulation of the N2 elicited by the presence of popout features irrespective of the relevance of the stimulus, has been observed only when participants actively searched for an item that differed from the others. It further confirms that the N2 is an attention-related rather than a bottom-up, perception-related component.

In the context of social neuroscience, relevant physical features, such as the race of a face, can modulate either the N1 together with both the N2 and the P2. White participants show larger N1 and P2 for Black faces (i.e., other-race) relative to White faces (i.e., own-race) and the opposite pattern on the N2 (for review, see Ito & Bartholow, 2009). Sheng and Han (2012) presented Chinese participants with own- and other-race individuals with either painful and neutral facial expressions, while

---

[8] The visual search paradigm is a kind of perceptual task that requires attention to actively scan the visual field in which a target stimulus is presented among a series of distractor stimuli. Participants are required to detect as fast as possible the target stimulus.

monitoring ERP responses during both a pain and a race[9] judgment tasks. Own-race faces, but not other-race faces with painful facial expressions elicited a more positive deflection in the P2–N2 range relative to neutral facial expressions, unravelling a racial bias in empathic response towards own- and other-race individuals' pain. A more recent study from the same laboratory (Sheng, Liu, Zhou, Zhou, & Han, 2013) demonstrated that such racial bias observed on the P2 component, is increased in participants treated with oxytocin, i.e. a neuropeptide involved in the development of trust among ingroup members.

The functional meaning of the N3 component remains instead still unclear.

As more clearly stated below, the above-mentioned ERP components have been associated with the mechanisms underlying the experience sharing because they appear to be modulated by physical facial features in a more automatic and bottom-up manner.

The P3 is, instead, one of the most studied ERP components in cognitive psychology, it is classically generated for rare and unexpected stimuli in an oddball paradigm but it is related also to the update of short-term memory, to the evaluation and categorization processes of the stimuli and as response to motivationally significant events (for a comprehensive review of the cognitive processes underlying the P3 elicitation, please see Polich, 2012).

In the context of the electrophysiological studies on empathy for pain, all the described components, either the earlier N1, P2, N2, N3 and the P3 are more positive for painful relative to neutral stimuli.

Importantly, as better described below, evidence suggests that the P3 is mainly related to mechanisms underlying the mentalizing because it appears to be modulated by the specific task participants are required to perform, in a top-down manner.

---

[9] In the race judgment task, participants are required to say whether the presented face belongs to either own or different race.

For instance, Decety and colleagues (Decety, Yang, & Cheng, 2010) provided early evidence about the involvement of down-regulation mechanisms on the empathic response to pain registered on the P3. In their study, 15 physicians and 15 control participants performed a pain decision task while their electrical activity was monitored. They were exposed to pictures depicting either painful or neutral stimulation. Whereas control participants showed an empathic response towards others' pain, as indexed by more positive P3 deflection in response to painful relative to neutral stimulation, physicians did not show such modulation (see Figure 5). Results of this study have been interpreted in terms of down-regulation of the response to pain in individuals working with patients. Likely, this down-regulation is adaptive, in order to have more cognitive resources available to be of assistance, and to consequently reduce personal distress.

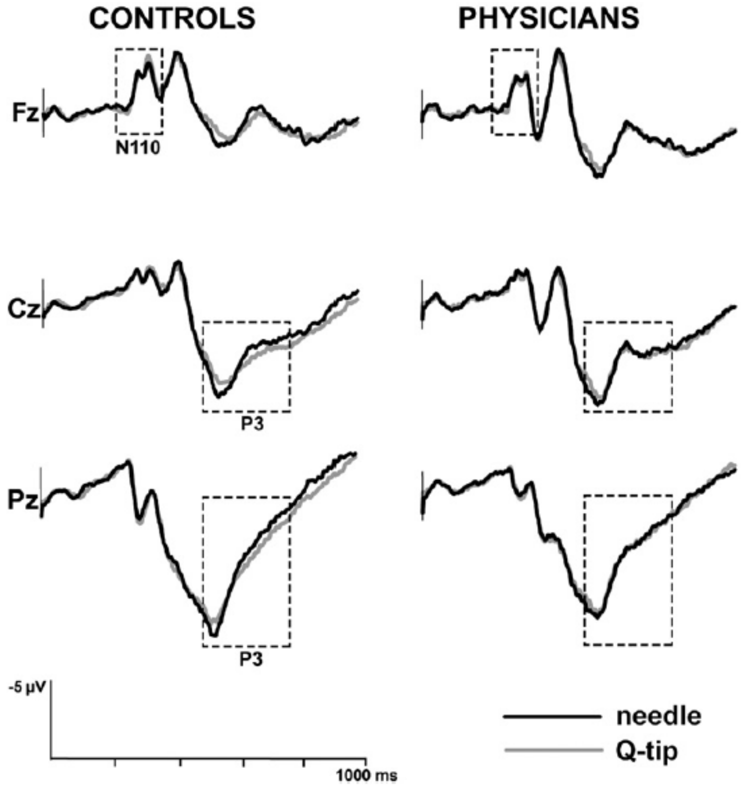


Figure 5. Results of Decety et al.'s study (2010). Control participants (left panel) showed larger (i.e., more positive) P3 for painful (i.e., body parts pricked by a needle) relative to neutral (i.e., body parts touched by a Q-tip), whereas physicians (right panel) did not.

Meng and colleagues (Meng, Hu, Shen, Yang, Chen, Huang, and Jackson, 2012) showed that although larger P3 amplitude is observed in response to painful relative to non-painful stimuli, such modulation can be increased if a negative, but not positive or neutral, emotional picture is presented as a prime before the target pictures depicting either painful or neutral stimulation, supporting the "threat value of pain hypothesis" (i.e., TVPH; Ibáñez, Hurtado, Lobos, Escobar, Trujillo, Baez, Huepe, Manes, Decety, 2011), see Figure 6. The TVPH posits that witnessing pain is potentially threatening for the observers and that might activate a threat-detection response instead of eliciting automatic empathic response.

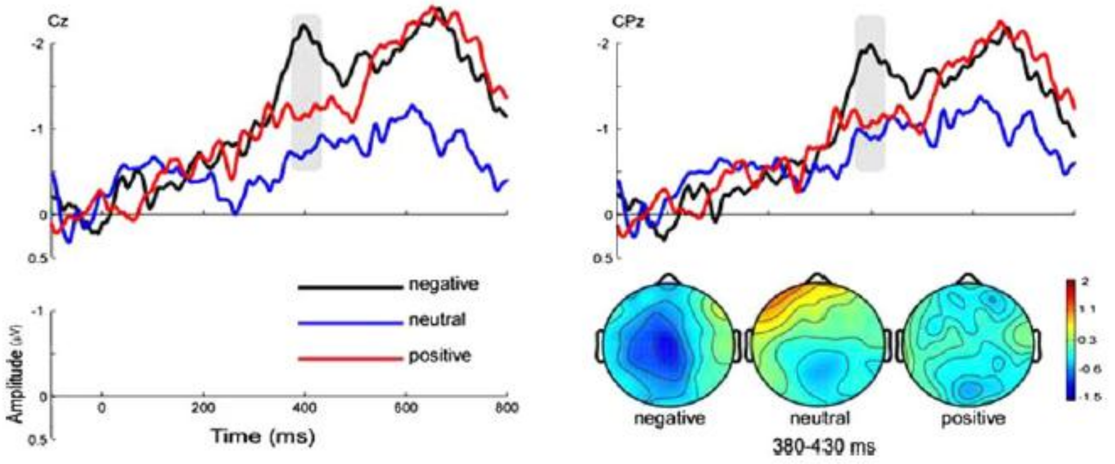


Figure 6. Differential P3 amplitudes (i.e., activity elicited in response to painful pictures minus activity elicited in response to non-painful pictures) in negative (black line), neutral (blue line), and positive (red line) prime conditions. The differential P3 in negative emotional prime condition is larger relative to neutral and positive prime conditions. The right bottom panel depicts the topographical maps of the scalp distribution of the differential waves in the time-range of the P3 in negative, neutral and positive emotional prime conditions.

Importantly, compelling evidence supporting the association between empathic response to others' pain and ERP components has been provided by a more recent study conducted in the same laboratory (Meng, Jackson, Chen, Hu, Yang, Su, and Huang, 2013). The authors directly investigated the interaction between pain perception in the self and observation of others by testing whether other-pain and self-pain primes differently modulate responsiveness to self-pain or other-pain targets in two ERP studies. Specifically, in one study participants underwent painful or non-painful heat stimulation after being primed with either painful or non-painful pictures and were required to rate the perceived pain. In the other study the self-pain or the self-heat stimulations were used as primes and stimuli depicting others' pain or non-pain were used as targets. Results showed that observing others' pain primes elicited faster reaction times, higher pain intensity rates and smaller P2 amplitude in response to self-pain stimulation; complementary, self-pain stimulation primes elicited faster reaction times, higher pain intensity rates and smaller P3 amplitude in response to others' pain. This pattern of results not only suggests that these ERP components are related to empathic responses but also supports the model of empathy that includes shared representations of others' pain.

The important contribute of electrophysiological research is that it can elucidate the time-course of empathy to other's pain.

Fan and Han (2008) registered electrical activity from Chinese adults required to perform a pain decision vs. a counting task about pictures of 1 or 2 hands painfully or not painfully stimulated. In the pain decision task, the authors observed that event-related potentials (i.e., ERPs) differentiated between painful and not painful conditions in a time-range from 140 ms until about 660 ms post-stimulus over the frontal-central area modulating a family of ERP components including P2, N2, N3 that appeared to be related to early stimulus-driven processes and the late P3 related to cognitive top-down processes. Thus, in the counting task, when attention was withdrawn from cues of pain, the effect of perceiving pain was intact on the early components (i.e., P2, N2 and N3) but significantly reduced on this later component, (i.e., P3) thus indicating − according to the well-known cognitive origin of the P3 (Donchin, 1981; Sessa, Luria, Verleger, &

Dell'Acqua, 2007) − that the cognitive processes involved in empathy needs attentional resources, see Figure 7.

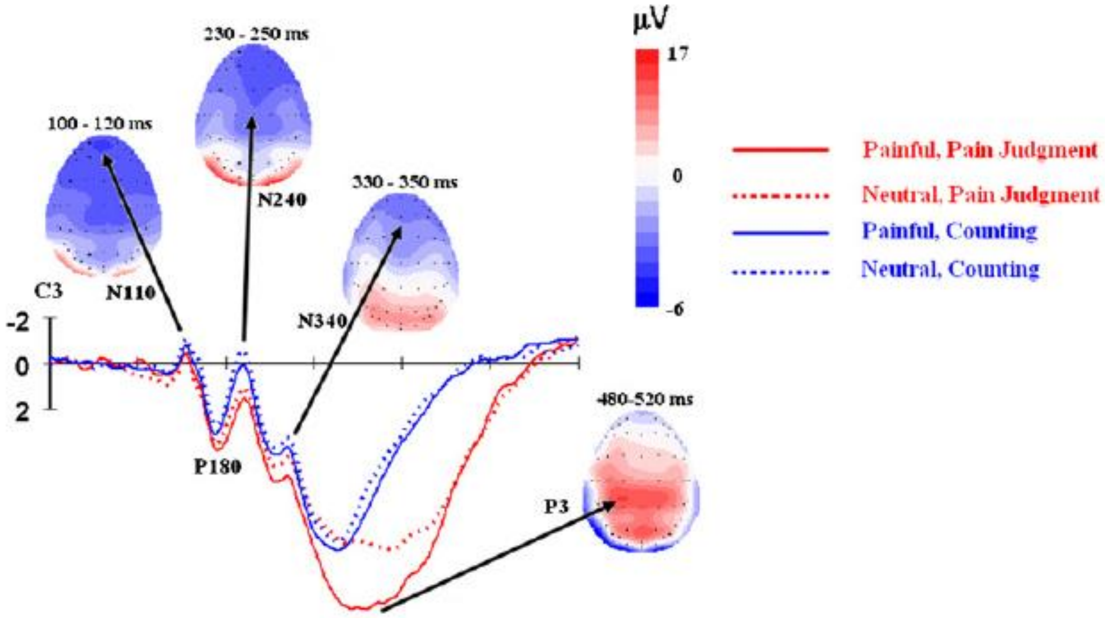


Figure 7. Results of Fan and Han's study. In the pain judgment task (red lines) a modulation of ERP components in response to painful stimuli relative to non-painful stimuli is observed between 140 and 660 ms. In the counting task (blue lines) such modulation is suppressed on the P3 component.

In similar conditions, Li and Han (2010) asked their participants to shift their perspective in performing the pain decision task from the self to that of an unfamiliar person. When taking other's perspective, ERP responses to pain were observed only in the time-range including N1, P2, N2 and N3 but not the P3, highlighting the dissociation between two mechanisms, one earlier, mediated in a bottom-up manner and thus related to stimulus-driven reactions that are task-independent (N1, P2, N2 and N3) and one later (P3) that is task-dependent and so related to cognitive processes (i.e., perspective shifting).

These results suggest, even though do not comprehensively clarify, that the early components could be associated with experience sharing processes whereas the P3 with mentalizing. The study proposed in Chapter 2: Dissecting empathy, had the main aim of addressing such dissociation in a direct way.

1.3 Variances in empathic processes

Empathy is a commonly experienced ability in healthy individuals, but extant evidence revealed a big variability in this ability both among individuals and as a function of contextual cues. Indeed, people may show differences in empathic response according to personality traits or might be influenced by either the contextual appraisal or specific characteristics of the person who is perceiving pain, such as race or social fairness.

Individual differences. For instance, Decety and Jackson (2004) highlighted that self-awareness is an important modulator of empathy, as an example of empathy variability as a function of individual differences. The capacity of recognizing one's own emotional state is at the basis of identification with others' emotional states (see also Asendorpf and Baudonnière, 1993). An empirical support to this view has been provided by Moriguchi and collegues (Moriguchi et al., 2007), which tested empathic abilities in healthy participants with high and low scores of alexithymia, which is a personality trait characterized by the difficulty in recognizing and expressing one's own emotional states. Alexithymic participants rated painful stimuli as less painful than nonalexithymic participants and showed lower scores in perspective taking and decreased neural activity in brain areas involved in empathy towards others' pain, such as DLPFC and ACC (for similar results, see also (Bird, Silani, Brindley, White, Frith, and Singer, 2010; Silani, Bird, Brindley, Singer, Frith, and Frith, 2008).

Reduced neural empathic response to others' pain has been observed also in children with conduct problems and callous traits (Lockwood, Sebastian, McCrory, Hyde, Gu, De Brito, and Viding, 2013) or in patients with congenital insensitivity to pain (Danziger, Prkachin, and Willer, 2006). On the contrary behavioral evidence suggests that people that report high scores on questionnaires assessing pain catastrophizing, i.e. a particular response to pain that includes elements of rumination, magnification and helplessness, perceive more intense pain in others (e.g. Sullivan, Martel, Tripp, Savard, and Crombez, 2006).

Of course, as already mentioned in the first section, people show interindividual variance in empathy traits. An easy way to assess empathy abilities is to administer participants with self-report empathy questionnaires, for instance the above described Interpersonal Reactivity Index (i.e. IRI; Davis, 1983) to either correlate the scores with neural responses or to establish the empathy range of participants sample, with for instance the Empathy Quotient (i.e. EQ) developed by Baron-Cohen and Wheelwright in 2004. In this self-report measure, there are 60 items including 20 filler items; responses are given on a 4-point scale ranging from 'strongly agree' to 'strongly disagree' and the EQ score may fall in different ranges: low empathy (i.e., 0–32), middle empathy (i.e., 33–52), high empathy (i.e., 53–63), extremely high empathy (i.e., 64–80) and maximum empathy (i.e., 80). An English and Italian version of the EQ is reported in Appendix 1 together with the IRI.

Several studies reported a significant association between hemodynamic response to others' pain and the cognitive perspective taking (e.g. (Cheng, Lin, Liu, Hsu, Lim, Hung, and Decety, 2007; Cheng, Yang, Lin, Lee, and Decety, 2008) or the affective empathic concern subscales of the IRI (e.g. Saarela et al., 2007; Singer et al., 2004; 2006). Further, participants with high scores in the cognitive fantasy subscale of the IRI showed greater sensorimotor contagion assessed by recording MEPs under TMS stimulation on participants' motor cortex (Avenanti et al., 2010).

Contextual cues. Race. In the study by Avenanti et al. (2010) another important modulator of empathy has been identified, namely the race of the person who perceives pain. Over the last decade, a large body of neuroimaging studies adopting variants of this approach have established that own-race and other-race faces are processed differently, likely at all stages of the identification process (e.g., Amodio, Harmon-Jones, & Devine, 2003; Cunningham, Nezlek, & Banaji, 2004; Hart et al., 2000; Wheeler & Fiske, 2005). Results from studies using event-related potentials (ERPs) are exemplar in this perspective. Race-driven modulations of face processing occur as early as the N1 component time-locked to face onset (Ito & Urland, 2003). Race, however, continues to modulate neural activity in cascade even during post-sensory stages of processing, up to and including working memory maintenance of face stimuli, as reflected in race-dependent effects observed on subsequent P2, N2, P3 (e.g., Dickter & Bartholow, 2007) and sustained posterior contralateral negativity (SPCN; e.g. Sessa, Tomelleri, Luria, Castelli, Reynolds, & Dell'Acqua, 2012) ERP components (see Ito & Bartholow, 2009, for a review).

Avenanti et al. (2010) used TMS to explore corticospinal excitability of both Black and White participants while they were exposed to short video-clips showing either a needle penetrating either own-race or other-race model’s hand (i.e. painful stimulation) or a Q-tip touching model’s hand (i.e. neutral stimulation), see Figure 8. MEPs were recorded to single-pulse TMS on the left motor cortex from both the corresponding muscle that participants observed being stimulating (i.e. the right FDI, first dorsal interosseus) and from the ADM, abductor digiti minimi, taken as control.

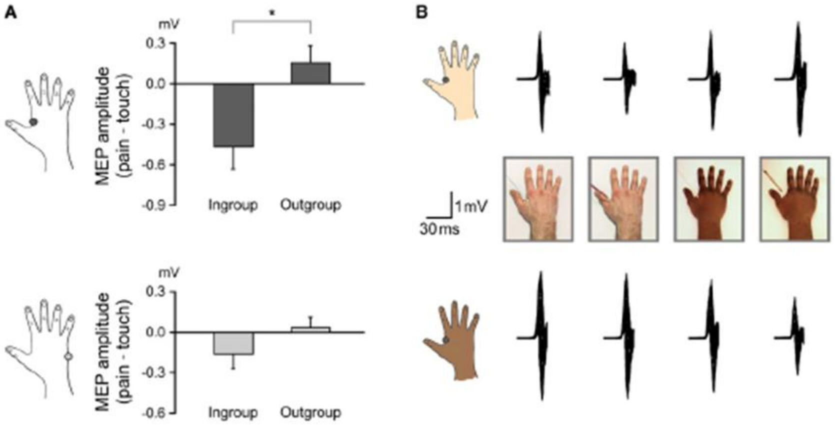


Figure 8. Results of Avenanti et al.'s study (2010). A) MEP amplitudes registered from the FDI (upper panel) and from ADM (lower panel) in response to to the painful, compared to the neutral, stimulation of both ingroup and outgroup hands. B) Example of experimental conditions.

The authors observed greater sensorimotor contagion for own- compared to other-race pain as revealed by greater FDI, but not ADM, corticospinal inhibition (i.e. similar to that observed as defense to self-experienced painful event) while observing own-race's pain. Interestingly, the corticospinal contagion for own-race targets was associated with participants' implicit racial prejudice towards other-race individuals – as assessed by the Implicit Association Test (i.e.; IAT Greenwald, Nosek, and Banaji, 2003). The IAT will be described in deeper detail in Chapter 3: Empathy towards other-race pain. Nonetheless, when both Black/White participants were presented with a model's hand of a different but unfamiliar race, i.e. a violet hand, greater sensorimotor contagion was restored, see Figure 9.

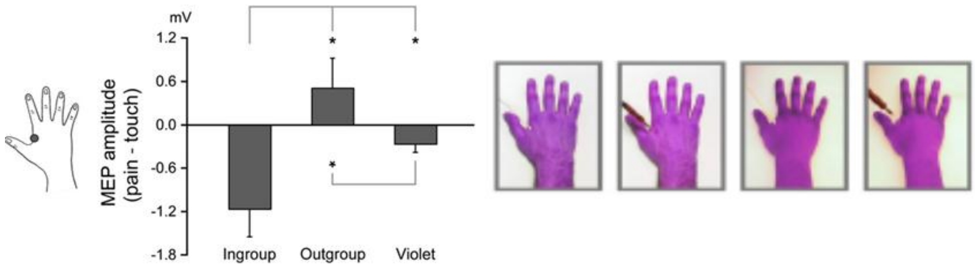


Figure 9. MEP amplitudes from the FDI in response to the painful, compared to the neutral, stimulation of violet hands.

These findings lead to an explanation of culturally learnt racial prejudice − that is race-specific −, which in turn modulates empathy-related brain responses (for similar results observed in pupil dilation see Azevedo et al., 2012; Forgiarini et al., 2011 for similar results observed in skin conductance response, SCR). However, an association between individual difference in implicit racial prejudice and cross-racial empathy for pain is not consistent with all the neuroscientific techniques, or it has not been addressed at all; I'll discuss such issue in deeper detail in Chapter 6: Conclusions.

Empirical support to the role of implicit race bias in the neural empathic response towards other-race pain has been provided also with fMRI by Xu, Zuo, Wang, and Han (2009) who observed analogous selective empathic response in Chinese and Caucasian participants towards own-race models. Similarly to Avenanti and collegues' study (2010), participants were exposed to short video clips featuring Chinese or Caucasian characters whose faces were either painfully (e.g., the face was penetrated by a needle) or neutrally (e.g., the face was touched by a Q-tip) stimulated. The participants' task was to categorize the video clips based on whether the characters were feeling pain or not, so-called pain decision task, disregarding their race. Bold-Oxygen-Level-Dependent (i.e., BOLD) responses recorded from the ACC was increased when participants watched faces under painful stimulation relative to faces stimulated with the non-painful object. Further, such BOLD response increase detected in ACC was of greater magnitude when the painful stimulation was applied to own-race characters compared to other-race characters. To note, results and conclusions from this study were compatible with the view proposed in prior work (Singer et al., 2004) that, among the diverse structures composing the pain matrix, the selective involvement of the ACC is a direct reflection of an empathic reaction of emotional/affective nature, which, in addition, depended on race (see also Sheng & Han, 2012), likely because the race of a face conveys any closer relationship between the observer and the target.

Race, as important element of variance in empathic processes, has been investigated in one of the proposed studies, which is described in Chapter 3: Empathy towards other-race faces.

Notably, physical facial features are processed in a bottom-up manner; however, in all of the cited studies the race of the target was completely task-irrelevant, so that it was implicitly appraised by participants.

Singer and colleagues (2006) have offered an elegant demonstration that also non-perceptual implicit information can modulate empathy, such as attribution of specific traits to perceivers. In the first phase of their study, the authors engaged male and female participants in an economic game (i.e., a sequential Prisoner’s Dilemma game) in which two confederates could play in either fair or unfair modality. In the second phase, BOLD fMRI signal was monitored in the same participants while a colored arrow signaled whether a painful stimulation would have been applied to either themselves (i.e., self condition), to the fair confederate (i.e., fair condition) or to the unfair confederate (i.e., unfair condition). Both female and males exhibited reduced activation of empathy-related brain regions (including AI, fronto-insular cortex and ACC) towards unfair confederates. Furthermore, in women this reduction in activity was small compared to men, which in turn showed hemodynamic response in some brain areas associated with desire of revenge, that is desire of punishment because of the social unfair behavior. These findings suggest that implicit appraisal of social fairness can modulate empathic neural responses.

Variance in empathic processes is complemented by the contribution of explicit information learnt by observers in the experimental context. Socially derogated targets (Harris and Fiske, 2006) or manipulated group membership constitute examples of explicit contextual appraisal (Eres and Molenberghs, 2013; Hein, Silani, Preuschoff, Batson, and Singer, 2010), so does inducing participants to believe that the painful stimulation is used as successful or useless therapeutic treatment (Lamm, Batson, and Decety, 2007) or feeling emotional pain for HIV/AIDS due to drug abuse patients relative to HIV/AIDS due to infected transfusion patients (Jean Decety, Echols, and Correll, 2009).

Trustworthiness. In current studies, implicitly perceived trustworthiness is proposed in Chapter 4 (Trustworthiness implicit appraisal) as additional modulator of empathy towards others' pain.

Indeed, people immediately form impressions of others on first meeting on the basis of others' physical appearance and immediately like or dislike them adjusting their behaviour even in the absence of previous knowledge on others' personality and social behavior. Evaluation of a stranger as trustworthy or untrustworthy is one of these appraisals 'at first sight' taking only a fraction of a second (e.g., Todorov, Said, Oosterhof, & Engell, 2011; Willis & Todorov, 2006). A large body of behavioral research has individuated physical characteristics that guide people in trustworthiness evaluation (e.g., Montepare & Dobish, 2003; Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008). Compelling advancement in this field has been provided by the work of Oosterhof and Todorov (2008) who first individuated facial features that people use to appraise others' trustworthiness: Faces appearing trustworthy are characterized by high inner eyebrows, pronounced cheekbones, wide chins and shallow nose sellion, while faces perceived as untrustworthy are characterized by low inner eyebrows, shallow cheekbones and thin chins and deep nose sellion; and secondly generated a database of standardized 2D models of trustworthy and untrustworthy faces, see Figure 10.

Figure 10. Oosterhof and Todorov's 2D models of trustworthy (right side of the continuum) and untrustworthy (left side of the continuum) faces.

Importantly, this appraisal seems reliable: Convincing evidence substantiates that individuals perceived as untrustworthy tend in fact to exploit the trust of others in social and economic exchanges (Stirrat & Perrett, 2010[10]); on the other hand, individuals perceived as trustworthy are more likely to possess a particular variation of the oxytocin receptor gene, known as the GG genotype, associated with a more prosocial and empathetic behavior (Kogan, Saslow, Impett, Oveis, Keltner & Saturn, 2011).

Trustworthiness of a face is a very intriguing characteristic because it is directly conveyed by the physical facial features, similarly to the race of a face; to note, it seems to be the physical counterpart of the social fairness.

I hypothesize that trustworthiness is a cue that people use at first sight to shape properly social interactions. In this vein, I hypothesize that even in absence of explicit information about either the social behavior or personality traits of the others, people

---

[10] Stirrat and Perrett's faces are not Oosterhof and Todorov's 2D models. They manipulated pictures of real faces in order to change the facial-width ratio, which was calculated as bizygomatic width divided by height, such as wide faces are more untrustworthy and slim faces more trustworthy.

might use such cue as a heuristics to either approach or avoid strangers, and that in turn might modulate empathy towards trustworthy and untrustworthy individuals.

1.4 Promises and pitfalls of neuroscience of empathy

At present, one of the main focus of research on empathy is whether experience sharing and mentalizing are isolated processes or rather deeply interactive.

In a previous paragraph, I elucidated evidence of an anatomical dissociation underlying these processing streams but nothing can still be argued about the functional dissociation. The principal obstacle in addressing such issue is mainly referred to a subtle logical fallacy in the great majority of the early studies on empathy: Usually tasks either engage experience sharing or mentalizing processes. First, it must be underlined that the involvement of one of them does not necessarily imply the concurrent involvement of the other just because of the experimental paradigm. This leads to a second fundamental factor: The experimental task, or the cue used in it, selectively activates either experience sharing or mentalizing. Dissociating *a priori* the possible contribution of the two prevents interaction between them, if present, from being uncovered.

This approach, indeed, contributes to build mutually exclusive single-process models of empathy and lead to two main consequences. First, it strengthens studies which support the predominant contribution of either experience sharing or mentalizing in human empathic abilities and with that, the partial or total theoretical exclusion of the other (see for example Brass, Schmitt, Spengler, & Gergely, 2007; Gallese et al., 2004; Hickok, 2008; Papeo, Corradi-Dell’Acqua, & Rumiati, 2011; Rizzolatti et al., 2009; Saxe, 2005). Second, it leads to an oversimplification of experimental tasks and cues aimed at investigating either one of the two processes. Commonly, neuroimaging studies aiming at isolating as much as possible hemodynamic response associated with experience sharing, remove elements useful to infer contextual information from pictorial sensorimotor cues; in fact, additional contextual information might enrich empathic experience in perceivers. This is usually the case of picture-based paradigms, described in the first section. On the contrary, mentalizing studies ask participants exclusively to appraise the context from descriptions of internal states or from colored cues signaling the internal state of someone who is not visible or, possibly, does not even exist. This is rather the case of cue-based paradigms.

Noteworthy, this approach is not negligible *per se* because it provided first steps in elucidating how experience sharing and mentalizing can be triggered on the way of empathy and allowed a first and rigorous exploration of how these processes are subjective to inter-individual and context-dependent variability.

Nonetheless, the importance of the use of more naturalistic paradigms in exploring social cognition needs to be taken in serious consideration. Indeed, oversimplifying experimental contexts do not directly test cognitive processes in realistic social contexts, which might shape in an unpredictable way brain activity deployed to perform the experimental task. Zaki and Ochsner (2009) highlighted that the processing of targets' states involves multimodal information that need to be dynamically integrated by perceivers. Further, external information is usually part of a social context that might constrain perceivers' interpretation of targets' internal states.

Crucially, empathy is an emergent property of empathic subprocesses that are engaged when one is dealing with others' internal states. This assumption implies that there is no clear evidence of how empathic subprocesses manifest behaviorally. Indeed, some scholars advocate that the definition of empathy as the ability to *share* leans on reverse inference. Thus, no brain-behavior relationship has been described to link overlapping brain areas involved in both first- and third-hand experience of pain. In addition, these brain areas, like the ACC and the insula, are largely involved in many other functions and so, at present, the relation between these brain regions and empathic processes is not univocal, since they could be activated by the experimental tasks to some unspecifiable extent.

To go beyond such limit, researchers have started searching a specific association between brain areas consistently involved in one empathic subprocess and some behavioral measure. An example has been provided by linking hemodynamic activation to prosocial/altruistic behavior. However, there is still no strict converging evidence about either which brain areas activation are predictive of altruism nor, by consequence of it, which empathic process needs to be mostly engaged in order to observe prosocial behavior. On one hand, Mathur, Harada, Lipke and Chiao (2010) suggested that altruistic motivation for one's own social group is predicted by greater activation in MPFC; Waytz, Zaki and Mitchell (2012) observed a key role of the DMPFC in predicting altruistic behavior, of course these regions have been exclusively associated with mentalizing (for similar results see also Telzer, Masten, Berkman, Lieberman, & Fuligni, 2011). On the other hand, Morelli, Rameson and Lieberman (2012) found that daily helping was associated with greater activity in the septal area, a region that has been only indirectly associated with maternal caregiving in humans (e.g., Lieberman, 2013). Nonetheless, recent behavioral study demonstrated that a good mediator of prosocial attitude in 18-months-old infants is rather mimicry (Carpenter, Uebel, & Tomasello, 2013). Edele, Dziobek and Keller (2013) observed that affective, but not cognitive, empathy explains altruistic behavior, assessed with monetary offers, in an economic game (i.e., the dictator game). In line with these findings, Cialdini, Brown, Lewis, Luce, and Neuberg (1997) suggested that prosocial behavior is due to emotional reaction towards others' states and aiming at reducing personal distress.

Of note, there is no evidence that empathy-related behavior is confined to altruism and prosocial attitudes. In addition, the leading corpus of research on empathy has mainly been expanded through fMRI technique.

Although its excellent spatial resolution allowed scholars to localize empathic subprocesses in the brain, its poor temporal resolution did not helped in disentangling crucial issues such as functional interplay between them.

In the light of pitfalls that have been listed, other techniques could enrich the knowledge about empathy.

The current studies have been conducted using the ERPs technique in order to study the lacking information about the time-course of experience sharing and mentalizing in the empathic response.

1.5 Overview of the studies

The following chapters consist in a review of five ERPs experiments I conducted, which deal with the multifaceted nature of empathy (Chapter 2: Dissecting empathy) and variances in empathic processes, as a function of others' race (Chapter 3: Empathy toward other-race faces) and others' perceived trustworthiness (Chapter 5: Empathy and Trustworthiness, Experiments 4 and 5) of a face, addressed through classical and modified versions of the pain decision task.

The pain decision task requires participants to decide whether presented stimuli receive either painful or neutral stimulation. Classically, stimuli are pictures depicting either faces or parts of the body while receiving one of the possible stimulations as illustrated in Figure 11.

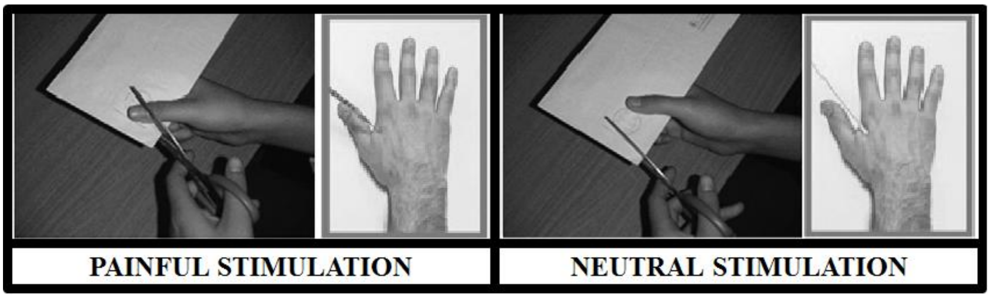


Figure 11. Example of stimuli used in empathy for pain research: pictures depicting either faces or part of the body while receiving either painful or neutral stimulation.

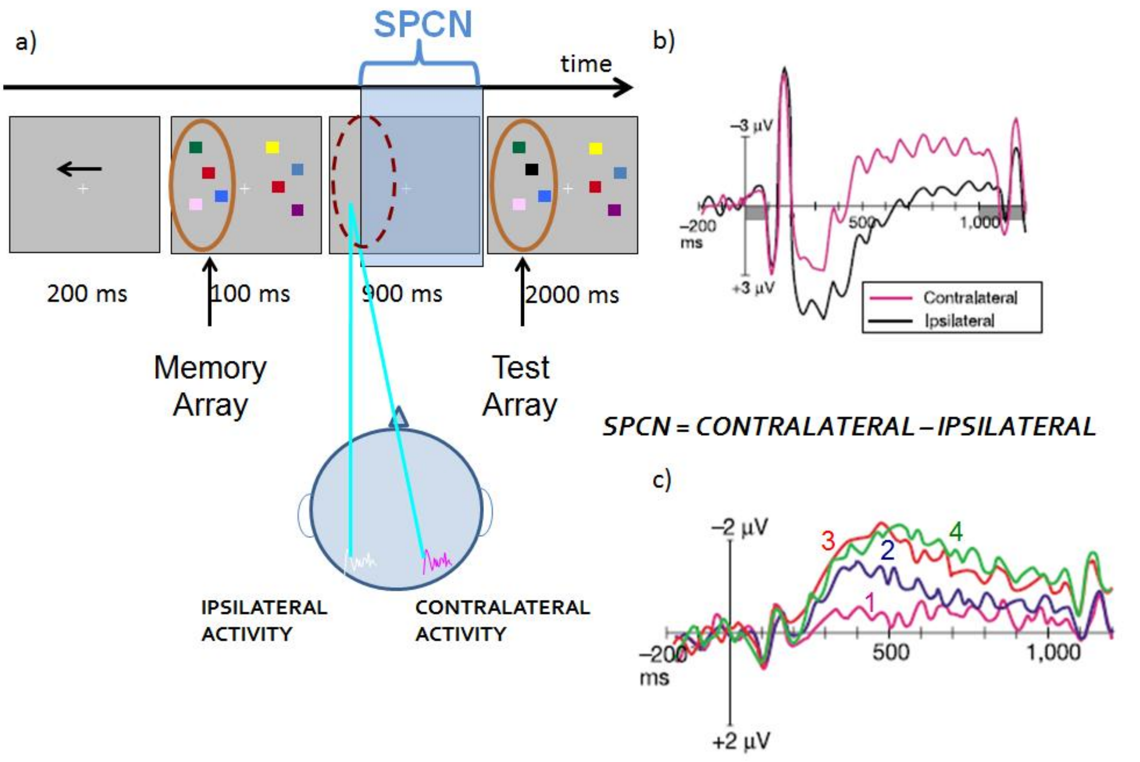
Furthermore, prior to investigate trustworthiness as modulator of neural empathic response, I adopted in Experiment 3 (Chapter 4: Trustworthiness implicit appraisal) a different paradigm, namely the change detection task, and a particular component of the ERPs, which is the Sustained Posterior Contralateral Negativity (i.e. SPCN), to test whether trustworthiness is automatically extracted from faces biasing visual working memory (i.e. VWM) processing.

The first experiment has arisen from the necessity to provide direct evidence, at least in the temporal domain, of the possible dissociation, or rather the interaction, between experience sharing and mentalizing. A first suggestion of a temporal dissociation between the empathic subprocess resulted from Fan and Han's study (2008). As I described in the second section of the current chapter, the authors observed that a family of early ERP components were modulated by painful context independently of the task requirements, indicating that in this early time-window ERP modulation is related to bottom-up mechanisms. By contrast, the P3 component was reduced when attention was withdrawn by painful context, as an index of down-regulation in the time-window including the P3.

In an attempt to make a step forward on the basis of the new flow of empathy research, I first propose a study in which both sensorimotor (picture of a face with either painful or neutral expression) and contextual information (a sentence describing either a painful or neutral context) are presented. This first experiment was aimed at testing directly the possible functional dissociation, or rather interaction, of the multifaceted nature of empathy in the temporal domain. Indeed, the temporal deployment of electrophysiological reaction could help in highlighting how the subprocesses functionally contribute at any given time.

Based on the results of this first experiment, I was able to dissect empathy into two time-windows, one earlier, pre-P3 associated with mechanisms underlying experience sharing, and one later, that includes the P3 component, associated with cognitive mentalizing. This is the starting point of the proposed studies on the way of exploring variance in neural empathic response. Neuroimaging and TMS studies suggested that people are more naturally empathic towards own-race individuals relative to other-race individuals (Avenanti et al., 2010; Xu et al., 2009). Noteworthy, such preference appeared to be confined to a particular subprocess of empathy, although the neuroscientific technique involved, i.e. fMRI and TMS, are suboptimal to clearly define this evidence. By virtue of excellent temporal resolution of the ERPs, I revealed that such implicit race bias is confined to experience sharing but empathic response towards other-race pain is observable in mentalizing time-window.

Although implicitly appraised, race of a face is processed quickly and automatically driven by physical facial features. Recently it has been demonstrated that evaluation of perceived individuals’ facial trustworthiness is appraised at first sight in a fraction of second (Willis & Todorov, 2006), similarly to race. I propose that trustworthiness plays another key role in modulating empathy even in the absence of previous knowledge on others’ personality and social behavior because implicit trustworthiness perception can quickly shape our social interactions. Oosterhof and Todorov (2008) individuated facial features that convey evaluation of trustworthiness perceived in others. In an attempt to determine the efficacy of such appraisal, I tested in Experiment 3 whether and how standardized physical facial features of trustworthiness, such as those of Oosterhof and Todorov’s computerized 2D models, are automatically maintained in VWM even when irrelevant for the task. VWM is a crucial step in defining the relevance of specific characteristics because it constantly feeds higher-order cognitive processes (e.g., long term memory) through the processing of lower-order functions (e.g., perception). To this aim, I adopted the change detection task, which is a paradigm specifically designed to test the resolution of VWM representations. Participants are required to memorize the identity of the face presented in the visual hemifield previously cued by an arrow and ignore the one presented in the non-cued hemifield; after a brief blank interval, the faces are presented again and participants have to say whether the face presented in the cued side was the same as that to memorize or not (for a schematic illustration of the early version of this task, see Figure 12a). I monitored the SPCN, which is an electrophysiological marker of the quantity (e.g., Vogel & Machizawa, 2004), the quality (e.g., Luria, Sessa, Gotler, Jolicoeur, & Dell'Acqua, 2010) and the resolution (e.g., Sessa et al., 2012) of visual information held in memory at any given moment during the retention interval, when the visual information was held in memory. The SPCN is a lateralized response characterized by greater negativity over the hemisphere contralateral to the visual hemifield where the target is presented relative to the response over the hemisphere ipsilateral to the target. To be visualized, the SPCN needs to be computed as the difference between contralateral and ipsilateral activity time-locked to the lateralized target stimulus (see Figure 12b,c).



Adapted from Vogel and Machizawa, 2004. *Nature*

Figure 12. a) Example of change detection task. In the panel is also depicted the time-window during which the SPCN is monitored. Participants are required to maintain in memory simuli presented in the visual hemifield previously cued by an arrow (i.e. colored squares in the red circle). After a retention interval, a test array is displayed and participants' task is to say whether the color of one of the to be held in memory targets has changed or not. The red circle has been drawn for representative aims. The SPCN is monitored during the retention interval and is the result of the difference between contralateral and ipsilateral hemispheric activities. b) Example of pre-subtracted activities in response to targets. c) Example of the modulation of SPCN (i.e. post-subtracted activity) as function of the number of to be held targets.

To anticipate, the SPCN trustworthiness is extracted from faces, even when task-irrelevant, modulating VWM representations of faces. Noteworthy, although implicit trustworthiness appraisal appears to be slower than explicit appraisal, its influence on higher-order cognitive processes, such as VWM, suggests that perceived trustworthiness might shape social interactions. Thus, I investigated empathy as a function of facial cues of trustworthiness.

In Experiment 4 I introduced, as first step, computerized Oosterhof and Todorov's trustworthy and untrustworthy faces in a modified version of the pain decision task. The modification consisted in temporally separating the face from the object inflicting painful stimulation because of the time needed to implicitly process trustworthiness. Finally, in order to use more ecologically valid stimuli, consistently with the new flow of research on empathy, I used real faces (Experiment 5) that I submitted to the rate of an independent sample of participants on the trustworthiness dimension. As expected, trustworthy faces induce magnified empathic reaction relative to untrustworthy faces.

Noteworthy, it has never been demonstrated that electrophysiological reaction to pain is empathy-specific. In fact, empathy is a multicomponential process and cognitive sub-mechanisms might occur. That would qualify the reaction to pain as a general activity elicited in some affective context. Indeed, pain decision task procedure might have overlapping mechanisms with those underlying an affective priming task. The affective priming is that effect occurring in a paradigm where two kinds of stimuli are presented sequentially and interleaved by a blank screen. To rule out the possibility that previous findings were not specific to empathic reactions to pain, I conducted Experiment 6 (reported in Appendix 2). I administered some participants with an affective priming task and I provided evidence supporting the empathy-specificity of the electrophysiological response observed in the presented studies.

Chapter 2

Experiment 1: Dissecting Empathy

In Experiment 1, I directly tested the hypothesis according to which the existing evidence of an anatomical dissociation between experience sharing and mentalizing are followed by complementing functional dissociation, or rather interaction, in the temporal domain.

As I discussed at great length in Chapter 1, the main focus of research on empathy is about its multifaceted nature and more specifically whether experience sharing and mentalizing are isolated processes or rather deeply interactive.

Although there is evidence of an anatomical dissociation underlying these processing streams, nothing can still be argued about the functional level. It must be noticed that the great majority of studies investigating empathy usually engage either experience sharing or mentalizing processes and in doing this they fall in a subtle logical fallacy at the interpretation level. First, it must be considered that the involvement of one of them does not necessarily imply the concurrent involvement of the other just because of the experimental paradigm. Most importantly, empathy subprocess, namely the experience sharing and the mentalizing, can be selectively activated by the specific experimental task, or by the cue used in it, and that might be taken as evidence that they can be activated independently. Certainly, dissociating *a priori* the possible contribution of the two prevents interaction between them, if present, from being unrevealed.

This approach, indeed, contributed to build mutually exclusive single-process models of empathy and led to two main consequences. First, it strengthened studies which support the predominant contribution of either experience sharing or mentalizing in human empathic abilities and with that, the partial or total theoretical exclusion of the other (see for example Gallese, Keysers & Rizzolatti, 2004; Hickok, 2008; Papeo et al., 2011; Rizzolatti et al., 2009; Saxe, 2005). Second, it led to an oversimplification of experimental tasks and cues aimed at investigating either one of the two processes. Commonly, neuroimaging studies aiming at isolating as much as possible hemodynamic response associated with either experience sharing or mentalizing, used commonly two kind of paradigms. Picture-based paradigms allow to enhance mechanisms underlying experience sharing because they focus on pictorial sensorimotor cues, on the other hand these paradigms totally miss contextual information, which might be crucial in experiencing empathy. By contrast, cue-based paradigms enhance mechanisms underlying mentalizing because they focus on contextual cues, by asking participants exclusively to appraise the context from descriptions of internal states or from colored cues signaling the internal state of the targets.

Consistently with this view, a recent meta-analysis on about 200 neuroimaging studies of human understanding of intentionality showed that at least mirror neuron and mentalizing systems are never concurrently active, each of them is specialized in the processing of either sensorimotor or contextual cues, respectively. According to this study, these systems are rather competitive and so mirror system does not aid the mentalizing in understanding others, nor viceversa (Van Overwalle & Baetens, 2009). However, there is some contrasting evidence, which suggests that they may cooperate during emotion understanding (e.g. Brass et al., 2007; Iacoboni, Lieberman, Knowlton, Molnar-Szakacs, Moritz, Throop, & Fiske, 2004; Rizzolatti & Craighero, 2004; Spunt, Falk, & Lieberman, 2010; Spunt, Satpute, & Lieberman, 2011).

Spunt and Lieberman (2012) explicitly tested such hypothesis and proposed a first integrative model of empathy by using an ecologically valid paradigm for the explicit identification and attribution of observed emotional expressions. They exposed participants undergoing fMRI to short silent video-clips taken from a dramatic television show, in which characters were mimicking a specific emotion. The two tasks where either to say how characters were showing their feeling (i.e., identification, *how*) or to infer why they were feeling it (i.e., attribution, *why*). Indeed, according to the authors, experience sharing and mentalizing could be functionally linked during attributional processing. Thus, once emotional expressions are identified (*how*) through the identification of motor acts in the face that allow specific expressions, such emotion can be attributed to an inferred social cause such as mental state (*why*). One advantage of this study is that stimuli were ecologically valid, that is sensorimotor and contextual cues were concurrently present at any given moment but the experimental tasks could strongly distinguish between the two systems. Results revealed dissociable yet functionally related contributes of mirror and mentalizing systems: Mirror, but not mentalizing, system was predominantly active in the explicit identification task; mentalizing, but not mirror, system, was predominantly active in the attribution task. However, during attributional processing the right posterior IFG (i.e. a brain area of the mirror system) activity precedes and was functionally related to the mentalizing system activity. These findings have been taken as evidence of an integrative model of mirror and mentalizing systems in understanding others' emotions.

Although the excellent spatial resolution of fMRI technique allowed scholars in localizing empathic subprocesses in the brain, its poor temporal resolution did not helped in deploying empathy abilities in the temporal domain so that it is still unclear *when*, if so, a functional interplay between them occurs.

A first suggestion of a temporal dissociation between the empathic subprocess resulted from an ERPs study by Fan and Han (2008). The authors observed that a family of early ERP components were modulated by painful context independently of the task requirments, indicating that in this early time-window ERP modulation is related to bottom-up mechanisms. By contrast, no modulation of the P3 component was found when attention was withdrawn by painful context, as an index of down-regulation in the time-window including the P3.

In an attempt to make a step forward on the basis of the new flow of empathy research, I conducted a study in which both sensorimotor (picture of a face with either painful or neutral expression) and contextual information (a sentence describing either a painful or neutral context in which the face is presented) are displayed. The temporal deployment of electrophysiological reaction could help in highlighting how the subprocesses functionally contribute at any given time.

2.1 Method

Participants

Data were collected from 20 volunteer healthy students (7 males) from the University of Padova (mean age: 25 years, SD = 3.14) who reported normal or corrected-to-normal vision and no history of neurological disorders. Only one was left-handed. All participants gave their informed consent according to the ethical principles approved by the University of Padova.

Stimuli and procedure

The stimuli were 32 sentences describing either a painful or neutral situation and 32 Caucasian face stimuli, 6 females and 6 males with either a neutral or painful expression, the same as those used in Sheng and Han (2012).

The face stimuli were scaled using an image-processing software so that each face fit in 2.9° x 3.6° (width x height) rectangle and the sentences were presented on three lines at the center of the screen in a 1.73° x 3.9° (width x height) virtual rectangle from a viewing distance of approximately 70 cm. Stimuli were presented on a 17-in cathode ray tube monitor controlled by a computer running E-prime software.

Participants performed a modified version of the pain decision task. Each trial began with the presentation of a fixation cross at the center of the screen, after 600 ms one of the 32 sentences (*contextual cue*) was displayed for 3000 ms at the center of the screen; after a blank interval of variable duration (800-1600 ms, jittered in step of 100 ms) one of the 32 faces (*perceptual cue*) was displayed for 250 ms. Participants were required to say whether the face had either a painful or neutral expression by pressing one of two response keys on the computer keyboard (“F” or “J”) using their left or right index finger, respectively, not before than 800 ms after the face offset. The mapping of the response keys was counterbalanced between participants. Each trial ended by asking participants to say the degree of their perceived empathy towards the presented face in the situation described on a 7-points Likert scale. This last display was needed to make sure the sentence to be considered and not simply ignored. The entire sequence of events is shown in Figure 1.

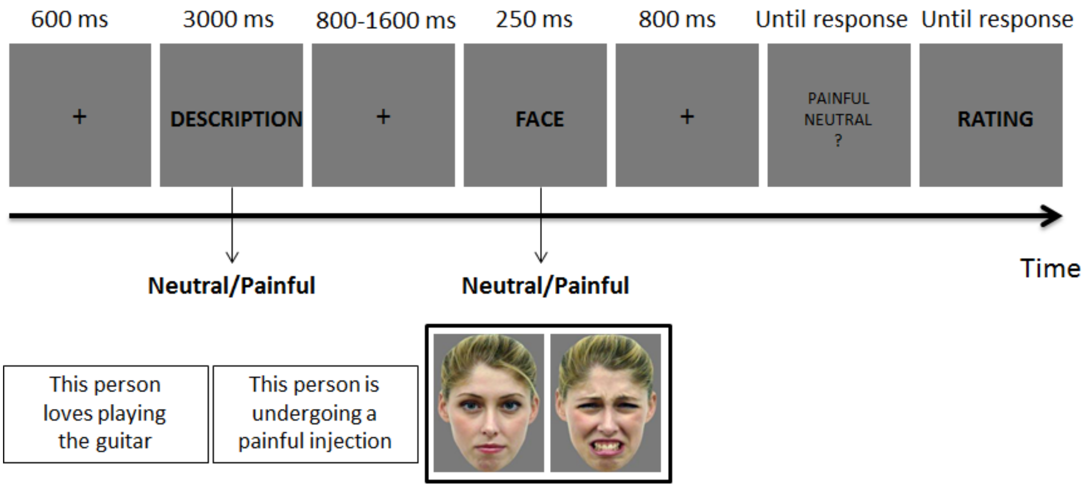


Figure 1. Example of the stimuli and the paradigm.

Faces and sentences were randomly intermixed for a total of 64 trials in each of 5 blocks.

The experiment lasted not more than 50 minutes. Before saving data participants performed a practical session in order to familiarize with the experiment and to learn to blink only after their response and to fix the gaze at the center of the screen.

EQ[11]. The EQ index was computed by adding points as follows: 2 points have been assigned when participants responded “strongly agree” and 1 point for “slightly agree” responses to the following items: 1, 6, 19, 22, 25, 26, 35,36, 38, 41, 42, 43, 44, 52, 54, 55, 57, 58, 59, 60. By contrast, 2 points have been assigned when participants responded “strongly disagree” and 1 point for “slightly disagree” responses to the following items: 4, 8, 10, 11, 12, 14, 15, 18, 21, 27, 28, 29, 32, 34, 37, 39, 46, 48, 49, 50.

---

[11] As the reader might remember from the third section of Chapter 1, the EQ has been developed by Baron-Cohen and Wheelwright (2004). In this self-report measure, there are 60 items including 20 filler items; responses are given on a 4-point scale ranging from ‘strongly agree’ to ‘strongly disagree’ and the EQ score may fall in different ranges: low empathy (i.e., 0–32), middle empathy (i.e., 33–52), high empathy (i.e., 53–63), extremely high empathy (i.e., 64–80) and maximum empathy (i.e., 80). See Appendix 1.

IRI[12]. The IRI does not have a total score, instead, each of its four subscales constitutes an index *per se*: PT, F, EC and DP scores. To calculate each index, reverse scores (i.e. assigning 5 when participants responded 1 and 4 when responded 2 and viceversa) in items 7 and 12 have been considered for the F subscale; in items 3 and 15 for PT subscale; in items 4, 14 and 18 for EC subscale and reverse scores in items 7 and 12 have been considered for PD subscale.

Electrophysiological recording and analyses

The EEG was recorded during the change detection task from 64 active electrodes distributed over the scalp in accordance with the international 10/20 system placed on an elastic Acti-Cap, referenced to the left earlobe. The EEG was re-referenced offline to the average of the left and right earlobes. Horizontal EOG (i.e., HEOG) was recorded bipolarly from two external electrodes positioned laterally to the left and right external canthy. Vertical EOG (i.e., VEOG) was recorded from Fp1 and one external electrode placed below the left eye. The electrode impedance was kept less than 10KΩ because of the highly viscous electro gel and the properties of active electrodes.

EEG, HEOG and VEOG signals was amplified (pass band 0.1-80 Hz) and digitized at a sampling rate of 250 Hz. The EEG was segmented into 1200-ms epochs starting 100 ms prior to the onset of the faces. The epochs were baseline corrected based on the mean activity during the 100-ms prestimulus period, for each electrode site. Trials associated to incorrect responses or contaminated by large horizontal eye movements, eye blinks or other artifacts (exceeding ± 30μV, ± 60μV and ± 80μV, respectively) were excluded from analysis. Separate average waveforms for each condition were then generated time-locked to the face. According to the visual inspection in butterfly

---

[12] As the reader might remember from the first section of Chapter 1, the IRI is a self-report questionnaire, which is composed of four 7-items subscales that measure both the capacity of taking others' perspective (Perspective Taking, i.e. PT, and Fantasy, i.e. F, subscales) and the capacity to feel with the others their feelings (Empathic Concern, i.e. EC, and Personal Distress, i.e. PD, subscales). Responses are given of 5-point scale ranging from 1 to 5. See Appendix 1.

view[13], I selected different time-windows as follows: 88-96 ms for the N1; 108-180 ms for the P2; 200-360 ms for the activity between N2-N3; and 400-840 ms for the P3. Electrical activity was explored over all the electrodes placed on the scalp.

Mean ERPs amplitude values were measured at pooled electrodes selected from the fronto-central (Fz, F1–F2, F3–F4, F5-F6, FCz, FC1–FC2, FC3–FC4, FC5-FC6), and centro-parietal (CPz, CP1–CP2, CP3–CP4, CP5-CP6, Pz, P1-P2, P3-P4, P5-P6) regions measured in time windows specifically selected for each ERP component, according to the visual inspection and previous work (e.g., Fan & Han, 2008). In all multi-factorial analyses, a Greenhouse-Geisser correction was used where appropriate.

Neural reactions to pain were calculated by subtracting mean amplitude values for neutral conditions (i.e. neutral information regarding either the facial expression, the description, or both of them) from those recorded for painful conditions. Three differential scores resulted from these computations. I'll refer to *picture-based empathic reaction* (i.e. *picture-based ER*) when neutral condition was subtracted from the condition in which only the facial expression was painful. I'll rather refer to *context-based empathic reaction* (i.e. *context-based ER*) when neutral condition was subtracted from the condition in which only the description was painful. I'll finally refer to *combined empathic reaction* (i.e., *combined ER*) when the neutral condition was subtracted from the condition in which both the facial expression and the description were painful.

At the end of the ERPs recording session, participants completed the two self-report measures of empathy: the IRI and the EQ.

---

[13] The butterfly view allows to overlap waveforms observed in each electrode for each condition. In this way, a very precise time window can be identified.

2.3 Results

Pain decision task. Reaction times (RTs) exceeding each individual mean RT in a given condition +/-2.5 *SD* and RTs associated with incorrect responses were excluded from the RT analysis. Individual mean proportions of correct responses and RTs were submitted to separate repeated measure analyses of variance (ANOVAs), both considering facial expression (painful vs. neutral) and description (painful vs. neutral) as within-subjects factors. Neither ANOVA showed significant factor main effects (max $F = 2.844$; min $p = .11$). Interestingly, individual mean proportions of correct responses showed marginally significant interaction between facial expression and description ($F = 3.670$; $p = .07$) in the direction of better accuracy for neutral (M=.99) relative to painful facial expression (M=.97) with neutral description and for painful (M=.99) relative to neutral facial expression (M=.98) with painful description, indexing a tendency to be more accurate in conditions of valence congruency.

EQ. Mean EQ score was similar to those found in the original study (Baron-Cohen and Wheelwright, 2004), i.e. 46.95/80 (*SD* = 9.92), so that the sample of participants reported an EQ score falling in the “middle empathy” range.

IRI. Inter-individual mean rating scores were: 3,31 (SD = 0,77) for F subscale; 3,79 (SD = 0,42) for PT; 3,824 (SD = 0,63) for CE; and 2,44 (SD = 0,61) for PD subscale.

ERPs: N1, P2, N2-N3 and P3. Figure 2 shows grand averages of the face-locked ERP components recorded at pooled fronto-central electrodes (upper panel) for each experimental condition in comparison to neutral condition (i.e. the condition in which both the face and the description were neutral in the same experimental trial). Noteworthy, a similar pattern of waveforms was observed at centro-parietal pooled electrodes (lower panel).

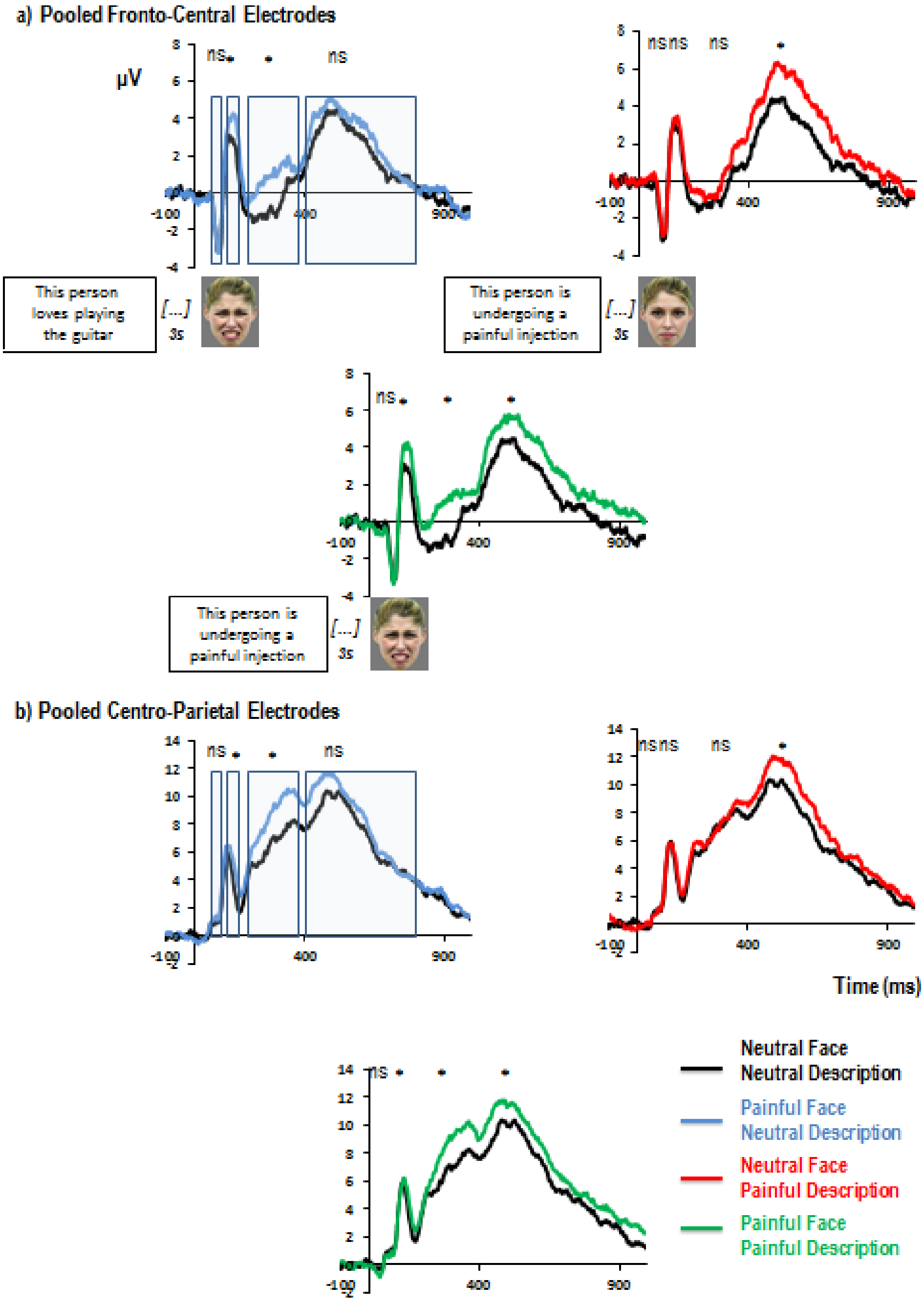


Figure 2. Grand averages of the face-locked ERP components recorded at pooled fronto-central and centro-parietal electrodes for each experimental condition in comparison to the neutral condition. a) The P2 and N2-N3, but not the P3, are significantly more positive in response to painful relative to neutral facial expression with neutral descriptions. b) The P3, but not the P2 and the N2-N3, is significantly more positive in response to painful relative to neutral description with neutral faces. c) The P2, N2-N3 and the P3 are significantly more positive when both the face and the description are painful relative to the neutral condition.

I submitted to ANOVA individual mean amplitude values of each component considering facial expression (painful vs. neutral) and description (painful vs. neutral) as within-subjects factors, separately for each region, i.e. fronto-central and centro-parietal pooled electrodes.

N1 (88-96). The ANOVA carried out on N1 mean amplitude values revealed no main effects over none of the pooled electrode regions nor interactions between facial expression and description (all $Fs < 1$). However, *context-based* and *combined ERs* to pain registered on the fronto-central pooled electrode sites, negatively correlated with the EQ score ($r = -424$ $p = .031$, for the *context-based ER*; $r = -391$ $p = .044$, for the *combined ER*) indicating that the higher was the EQ score the little was the empathic reaction.

P2 (108-156). The ANOVA carried out on P2 mean amplitude values revealed main effect of facial expression at pooled fronto-central sites $F(1,19) = 8.269$, $p = .01$, $\eta_p^2 = .303$, and a marginally significant effect at centro-parietal pooled electrodes $F(1,19) = 3.609$, $p = .07$, $\eta_p^2 = .160$), indicating that painful facial expressions elicited larger P2 (at fronto-central pooled electrodes 3.419 μV; at centro-parietal electrodes 5.393 μV) than neutral facial expression (at fronto-central pooled electrodes 2.559 μV; at centro-parietal electrodes 4.736 μV) irrespective of the description. The main effect of description did not reach significance level nor did the interaction between factors (all $ps > .09$).

Planned comparisons revealed that the facial expression elicited larger P2 amplitude than neutral face at both fronto-central and centro-parietal pooled electrodes (all $ps < .005$) and that the condition in which both facial expression and the description were painful was not significantly different to the conditions in which either only the face or the description was painful (all $ps > .05$).

In addition, I further explored the functional significance of the P2 reaction to pain in relation to empathy subprocesses. Fronto-central pooled electrodes showed a negative correlation between individual *picture-based ER* to pain and the PT subscale scores of the IRI, $r = -.545$ $p = .007$ indicating that the higher was the PT score the smaller was the *picture-based* ER on this ERP component at the fronto-central area. That might be due to the fact that on this early time-window, people with high cognitive empathy abilities need contextual information, lacking in this specific condition, to trigger empathic reaction. By contrast, indeed, positive correlations have been observed at centro-parietal pool between individual *context-based ER* to pain and the EC subscale of the IRI: $r = .386$ $p = .046$ indicating that the higher was the EC score the greater was the empathic reaction to painful description on this ERP component at centro-parietal area. That might be an index of the fact that people with high empathy abilities related to experience sharing are more efficient in triggering an empathic reaction in this early time-window even conveyed only by contextual information.

N2-N3 (200-360). The ANOVA carried out on N2-N3 mean amplitude values revealed the main effect of facial expression at both fronto-central pooled electrodes, $F(1,19) = 12.875$, $p = .002$, $\eta_p^2 = .404$, and centro-parietal pooled electrodes $F(1,19) = 19.656$, $p < .000$, $\eta_p^2 = .508$. The main effect of description did not reach significance level ($F < 1$) nor did the interaction between factors ($F(1, 19) = 1.261$, $p > .05$).

Planned comparisons revealed that the condition in which both facial expression and the description were painful was not significantly different from the conditions in which only the face was painful (all *ps* > .05) but was significantly different to the conditions in which only the description was painful ($t = 3.423$, $p = .003$) at centro-parietal pooled electrodes indicating that in this time-window the presence of painful facial expression does contribute in enhancing empathic reaction to others’ pain.

In addition, I further explored the functional significance of the N2-N3 reaction to pain in relation to empathy subprocesses. Interestingly, in this time-range, N2-N3 reaction to pain was associated with individual differences in EC scores at both fronto-central and centro-parietal electrodes. Figure 3 shows positive correlations between EC scores and individual both *picture-based* (*r* = .432 *p* = .028 at fronto-central area; *r* = .451 *p* = .023 at centro-parietal area) and *combined ERs* to pain and the (*r* = .357 *p* = .061 at fronto-central area; *r* = .453 *p* = .022 at centro-parietal area) indicating that the higher was the EC score the greater was the empathic reaction to either painful facial expression and neutral description or both painful facial expression and description in this time-range. This strongly suggests that the N2-N3 empathic reaction is strongly related to a subprocess of empathy associated with experience sharing mechanisms.

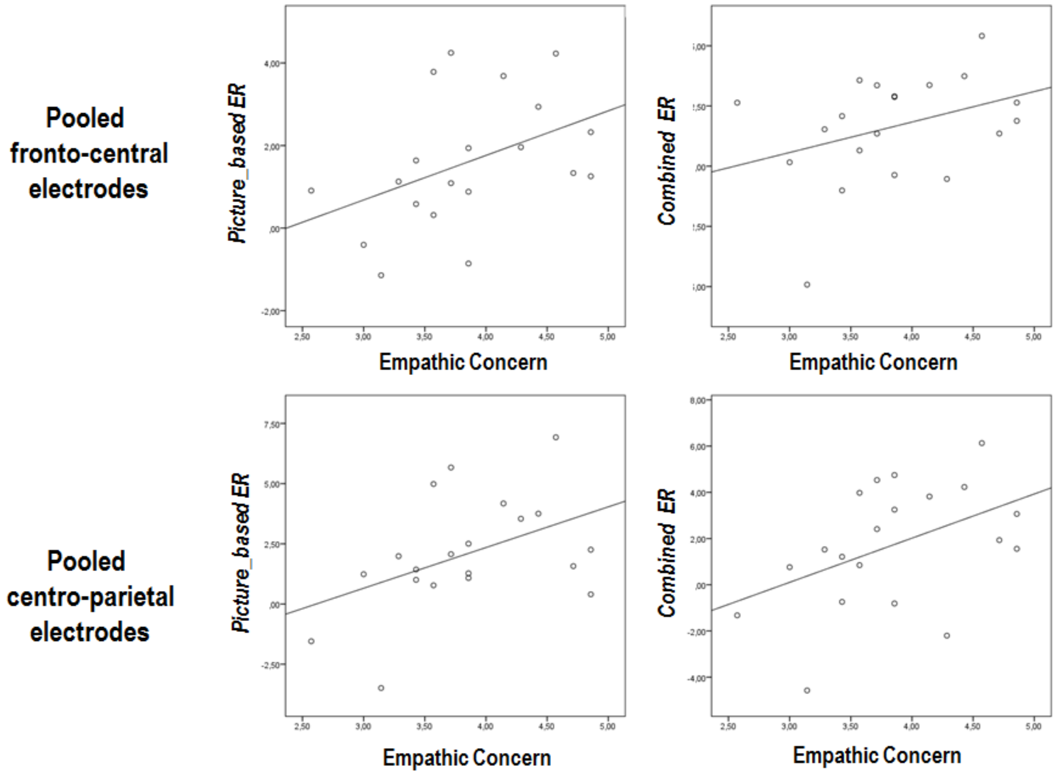


Figure 3. Scatterplots of correlations between individual empathic reactions (i.e. *picture based*, left panel; *combined*, right panel) and EC scores, registered at pooled both fronto-central and centro-parietal electrodes.

P3 (400-840). The ANOVA carried out on P3 mean amplitude values revealed main effect of description at both fronto-central pooled electrodes, $F(1,19) = 8.504$ $p = .009$ $\eta_p^2 = .309$, and at centro-parietal pooled electrodes $F(1,19) = 4.665$ $p = .044$ $\eta_p^2 = .197$. The main effect of facial expression did not reach significance level nor did the interaction between factors (all $ps > .05$).

Planned comparisons revealed that the condition in which both facial expression and the description were painful was not significantly different from the conditions in which either only the description was painful (all $ps > .05$).

As I did for the other ERP components, I further explored the functional significance of the P3 component in relation to empathy subprocesses. Interestingly, the fronto-central electrodes showed marginally significant negative correlation between individual *picture-based ERs* to pain and the PT scores: $r = .370$ $p = .054$ indicating that the higher was the PT score the little was the empathic reaction to either painful facial expression and neutral description on the P3. That further confirms correlations found on the P2. Thus, people with high cognitive empathy abilities need contextual information to trigger an empathic reaction.

Hence, results are clear-cut: in the time-window including the P2 and the N2-N3 the *perceptual cue* selectively activates mechanisms underlying the experience sharing; in the immediately following time-window, which includes the P3, the *contextual cue* selectively activates mechanisms underlying the mentalizing. This selectivity suggests that experience sharing and mentalizing are dissociable.

However, the visual inspection of the waveforms suggests that the time-window between the N2-N3 and ascendant part of the P3 may represent a critical time-window, in which processes involved can act as a hub between the experience sharing and mentalizing. Aiming at exploring this possible connection, I selected an additional time-window, i.e. 380-520 ms between the N2-N3 and the P3.

N2-N3–P3 (380-520). The ANOVA carried out on mean amplitude values of this time-window revealed a main effect of facial expression at centro-parietal pooled electrodes ($F(1,19) = 5.699$, $p = .028$, $\eta_p^2 = .231$), a main effect of description at fronto-central pooled electrodes ($F(1,19) = 6.292$, $p = .021$, $\eta_p^2 = .249$). Most importantly, interaction between the factors was observed at both fronto-central ($F(1,19) = 4.414$, $p = .049$, $\eta_p^2 = .189$) and at centro-parietal ($F(1,19) = 4.942$, $p = .039$, $\eta_p^2 = .206$) pooled electrodes. Post-hoc analysis showed that neutral condition was significantly more negative than conditions in which either only the description, the facial expression or both of them were painful (all $ps < .05$). In addition, at fronto-central pooled electrodes the comparison between the neutral condition and the condition in which only the facial expression was painful was only marginally significant ($t = 1.805$, $p = .08$)[14].

These findings are taken as evidence that there is an intermediate time-window the face might be processed at the level of mentalizing because participants are required to rate their empathy towards the face's pain.

2.3 Discussion

The results of Experiment 1 showed that experience sharing and mentalizing are two dissociable mechanisms. In the time-window including the P2 and the N2-N3 the *perceptual cue* selectively activated mechanisms underlying the experience sharing; in the immediately following time-window, which includes the P3, the *contextual cue* selectively activated mechanisms underlying the mentalizing.

---

[14] For the sake of completeness, I report that at fronto-central pooled electrodes a negative correlation between individual picture-based ERs to pain and the F scores was observed ($r = -.402$, $p = .039$) indicating that the higher was the F score the smaller was the empathic reaction to painful facial expression and neutral description in this restricted time-window relative to the P3; centro-parietal pooled electrodes showed, instead, a negative correlation between individual picture-based ERs to pain and the EQ scores: $r = -.399$ $p = .041$ indicating that the higher was the EQ score the smaller was the empathic reaction to painful facial expression and neutral description in this time-window. That further confirms correlations found on the P2. Thus, people with high cognitive empathy abilities need contextual information to trigger an empathic reaction.

However, a more fine-grained inspection of the waveforms suggested that there was an intermediate time-window between N2-N3 activity and the P3 peak sensitive to both cues of pain.

Crucially, planned comparisons revealed that the concurrent presence of both painful facial expression and painful description did not enhance the empathic reaction to pain. Indeed, neither the P2, the N2-N3 or the P3 components were significantly more positive in this condition relative to the conditions in which only one of the cues was conveying painful information. Moreover, empathic reaction to pain observed in these time-windows have been associated with specific empathic abilities, underlying either experience sharing or mentalizing.

Consistently with Fan and Han’ s first suggestion (2008), I demonstrated that the time-window which precedes the P3 peak, namely including the N2-N3 activity, is related to experience sharing: *picture-based* and *combined ER* were larger the higher was the score of participants obtained in the EC subscale of the IRI, which directly measures experience sharing abilities in empathy. By contrast, the time-window including the P3 was related to mentalizing: *picture-based ER* was reduced the higher was the score of participants obtained in the PT subscale of the IRI, which measures cognitive abilities underlying mentalizing. Indeed, such correlation suggests that painful information conveyed by the sensorimotor cue is not sufficient to trigger an empathic response in those participants with high cognitive abilities.

Interestingly, similarly to the P3, the P2 time-window revealed that *picture-based ER* was negatively correlated with PT scores at the fronto-central region. That might be due to the fact that on this early time-window, people with high cognitive empathy abilities need contextual information, lacking in this specific condition, to trigger empathic reaction. At the same time, *context-based ER* positively correlated with the EC scores at the centro-parietal area: the higher was the EC score the greater was the empathic reaction to painful description and neutral facial expression. That might be an index of the fact that people with high empathy abilities related to experience sharing are more efficient in triggering an empathic reaction in this early time-window even conveyed only by contextual information.

Intriguingly, these findings are in line with a recent study by Sheng and Han (2012). The authors presented participants with suffering and neutral facial expressions, while monitoring ERP responses during a pain judgment task. Painful facial expressions elicited ERP responses that were characterized by a more positive deflection in the P2–N2 range relative to neutral facial expressions. The potential source of this ERP reaction to pain in the P2 time window was the ACC. Furthermore, P2 amplitude values in subtracted face-locked ERPs (i.e., ERPs elicited by painful facial expressions minus ERPs elicited by neutral facial expressions) correlated positively with EC scores of the IRI.

Altogether, these results appear to indicate that the P2 component is strictly related to mechanisms underlying the experience sharing.

So, in line with the meta-analysis conducted by Van Overwalle and Baetens (2009), results of Experiment 1 showed that experience sharing and mentalizing are extensively dissociable mechanisms and that experience sharing does not facilitate mentalizing mechanisms.

However, there is an intermediate time-window between the N2-N3 and the P3 peak ssensitive to both as indicated by the effect of the interaction between facial expression and description registered in this time-range (i.e., 380-520). Indeed the neutral condition was significantly more negative than conditions in which either only the description, the facial expression or both of them were painful.

Spunt and Lieberman (2012) provided evidence of particular ecologically valid case in which mirror neuron system precedes and can be functionally related to the mentalizing system and proposed on the basis of their findings an integrative model of empathy. According to this model the mirror systems facilitates the rapid identification of facial expressions, which is then used as inputs to attributional processing in the mentalizing system. Furthermore, Brass et al. (2007) suggested that mirror system can be involved in action understanding in those cases in which familiar motor acts are observed in congruent and usual contexts (for a review, please see also Rizzolatti & Craighero, 2004). The results of Experiment 1 nicely dovetail with these previous findings by suggesting that although experience sharing and mentalizing are extensively dissociable, there is an intermediate time-window that temporarily falls between these two subprocesses during which they appear to be, to a certain degree, functionally connected. At the light of Spunt and Lieberman's findings, I speculate that this critical time-window acts as a node of transfer of information from experience sharing to mentalizing processing in the comprehension of others' pain.

Chapter 3

Experiment 2: Empathy towards other-race pain

In Experiment 2, the influence of the race of a face was investigated as modulator of neural empathic response.

Over the last decade, a large body of neuroimaging studies adopting variants of this approach have established that own-race and other-race faces are processed differently, likely at all stages of the identification process (e.g., Amodio et al., 2003; Cunningham et al., 2004; Hart et al., 2000; Wheeler & Fiske, 2005). Results from studies using event-related potentials (ERPs) are exemplar in this perspective, by revealing that race influence can be related to perception (Ofan, Rubin, & Amodio, 2011), attention (Amodio, 2010; Dickter & Bartholow, 2007; Ito & Urland, 2003) or maintenance of other-race faces in visual working memory (Sessa et al., 2012). Indeed, race-driven modulations of face processing occur as early as the N1 component time-locked to face onset and its effects are reflected up to subsequent P2, N2, P3 and Sustained Posterior Contralateral Negativity (i.e.,SPCN) ERP components (see Ito & Bartholow, 2009, for a review).

In the introduction section, I reported that recent neuroimaging and TMS studies revealed that race can also bias the processing of elements other than those strictly necessary to build and maintain a visual representation of a face: The ability to empathize with other persons, which appear to be selective towards own-race individuals when compared to that towards other-race individuals (Avenanti et al., 2010; Xu et al., 2009). Noteworthy, such preference appeared to be confined to a particular subprocess of empathy, although the time-scale of neuroscientific technique involved, i.e. fMRI and TMS, are suboptimal to capture effects arising from subtle factorial interactions and so clearly define this evidence. Indeed, although both studies provided fundamental information about *where* the potential neural loci of the interplay between racial bias and empathy may be localized in the brain, these studies are virtually tacit relative to *when* such interplay takes place

In both Xu's et al. (2009) and Avenanti's et al. (2010) studies, participants were exposed to face/hand stimuli conveying racial and empathy-eliciting information for a relatively long interval prior to recording their joint effects on stimuli processing. The use of long stimuli exposure makes it plausible that the affective/emotional and sensory-motor reactions described in those studies were in fact both mediated by higher level processes linked to the cognitive aspects of empathy, including attention, emotion regulation and cognitive evaluation of others' pain. In this vein, the long stimuli exposure provided participants with the opportunity to create a cognitive representation of others' suffering which may have in turn mediated — in a top-down fashion, as hinted by Decety and Jackson (2004; 2006; see also Decety, 2011)) — their emotional/affective and sensory-motor reactivity to the stimuli used by Xu et al. (2009) and Avenanti et al. (2010). Fan and Han (2008) described the time-course of the neural empathic response by suggesting a temporal dissociation between bottom-up and top-down mechanisms, the firsts in a time-window preceding the P3 component and the seconds in a time-window including the P3 component. In addition, Experiment 1 provided direct evidence of extensive functional dissociation in the temporal domain of experience sharing and mentalizing associated with these two time-windows. Mapping out the temporal locus of the interplay between race and empathy-related subprocesses, was the primary scope of Experiment 2, as it would allow determining the nature of the influence of race on empathy.

White participants in the present ERP study were presented with own- (i.e., White) and other-race (i.e., Black) faces in a painful condition (i.e., penetrated by a needle) or in a non-painful condition (i.e., touched by a Q-tip) under the requirement to perform a pain decision task. Extant evidence of modulations of ERP responses to faces as a function of race occurring as early as the N1 component would lead to hypothesize that, if an interaction between race and pain stimulation conditions were evident in ERPs, it might surface relatively early in time, within a pre-P3 onset time window, thereby suggesting that neural resonance is reduced when facing other-race individuals in pain. An alternative scenario would be more congruent with our interpretation of Xu's et al. (2008) and Avenanti's et al. (2010) results, namely, that of a cognitive involvement in the reduction of brain responses to the pain of other-race individuals reflected in interactive effects largely confined to the P3 time-range.

Noteworthy, results from Experiment 1 suggest that the racial bias observed by Xu et al., and Avenanti' et al. might be related to experience sharing and so to the pre-P3 rime-range. In this vein, in order to better qualify the meaning of each ERP component modulated by pain and/or race, participants in the present study were also administered, at the end of the pain decision task, a standard-race IAT and the Italian version of the IRI (Italian version by Albiero, Ingoglia, & Lo Coco, 2006). These measures were collected in order to assess parametrically whether implicit racial bias, as assessed by the IAT, influenced ERP reactions to pain, and whether affective and cognitive components of empathy, as assessed by the IRI, correlated with earlier (i.e., pre-P3) and later (P3 and following) ERP reactions to pain.

3.1 Method

Participants

Informed consent was obtained from twelve White students at the University of Padova (8 females; mean age = 26.4 years, *SD* = 8.6; 1 left-handed) volunteered to participate in the present experiment.

Stimuli and procedure (ERP recording session). The stimuli were colored digital photographs of 20 White (i.e., own-race; 10 males and 10 females) and 20 Black (i.e., other-race; 10 males and 10 females) faces with a neutral expression selected from Minear and Park's (2004) database. Each photograph was scaled to fit in a rectangular portion of the computer screen subtending 3.7° x 3.1° (width x height) of visual angle at a viewing distance of approximately 70 cm. Each face was manipulated digitally so as to be displayed in two different conditions. In the painful stimulation condition, the face was displayed with a needle of a syringe penetrating the cheek. In the non-painful stimulation condition, the face was displayed with a Q-tip touching the cheek. Figure 1a shows a sample of the stimuli (in grayscale), one for each experimental condition.

Each trial began with a fixation cross at the center of the computer screen that remained visible for a randomly variable duration (800–1600 ms, jittered in 100 ms steps). The fixation cross was then replaced by a centrally presented face, which was displayed for a fixed duration of 250 ms, and followed by a blank interval that lasted until response detection (see below). Figure 1b illustrates the temporal structure of one trial.

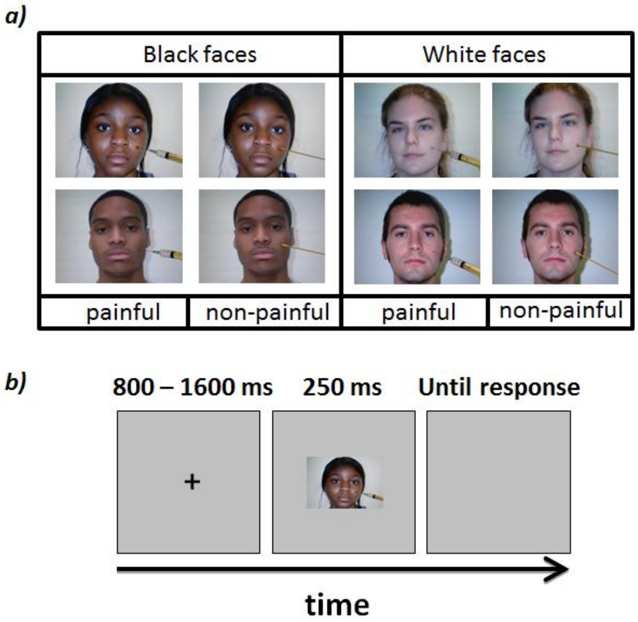
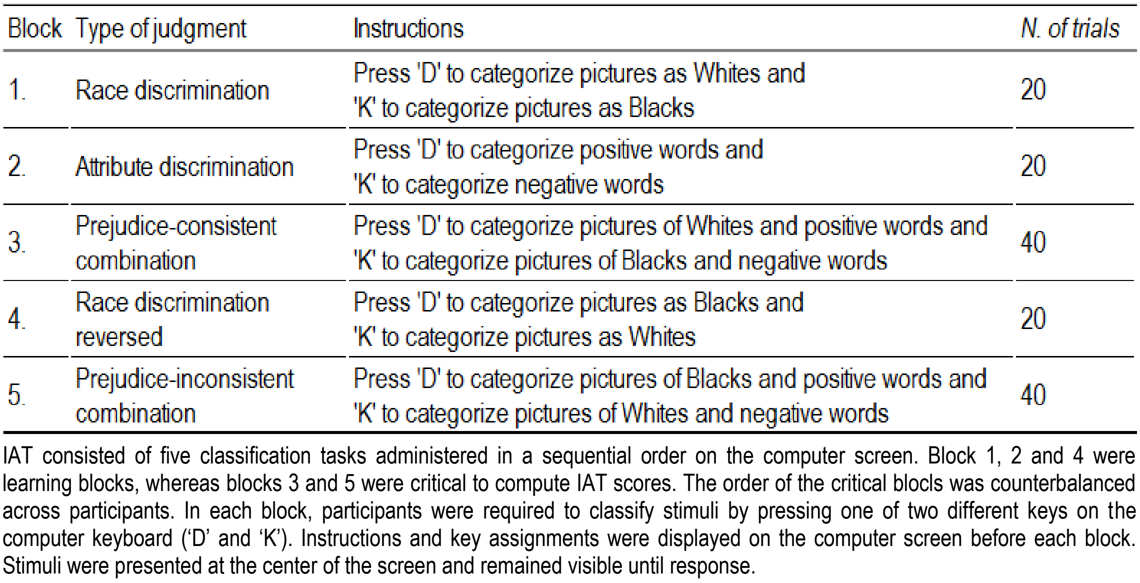


Figure 1. a) Examples of stimuli, one for each experimental condition: stimulation condition (painful vs. non-painful) and race (White/own-race faces vs. Black/other-race faces). b) Schematic illustration of the experimental procedure.

Participants were required to decide whether each face was painfully or not painfully stimulated by pressing one of two appropriately labeled keys of the computer keyboard (‘F’ or ‘J’, counterbalanced across participants) as quickly and accurately as possible. The experimental list of stimuli was organized in 12 blocks of 80 trials, which were preceded by a brief session of practice with stimuli that were not included in the set of experimental stimuli. In each block, all possible combinations of face’s sex, race, and stimulation condition were equally represented and randomly ordered at run-time for each participant.

IAT. The stimuli were colored digital photographs of 10 White male and 10 Black male faces that were different than the faces used for the ERPs recording session. The attributes were 10 positive words (e.g., peace, love, and paradise) and 10 negative words (e.g., disgust, vomit, and earthquake). A standard race-IAT was administered to each participant following the pain decision task. A description of the IAT procedure is illustrated in Table 1.

Table 1. Schematic description of the IAT procedure



Block	Type of judgment	Instructions	*N. of trials*
1.	Race discrimination	Press 'D' to categorize pictures as Whites and 'K' to categorize pictures as Blacks	20
2.	Attribute discrimination	Press 'D' to categorize positive words and 'K' to categorize negative words	20
3.	Prejudice-consistent combination	Press 'D' to categorize pictures of Whites and positive words and 'K' to categorize pictures of Blacks and negative words	40
4.	Race discrimination reversed	Press 'D' to categorize pictures as Blacks and 'K' to categorize pictures as Whites	20
5.	Prejudice-inconsistent combination	Press 'D' to categorize pictures of Blacks and positive words and 'K' to categorize pictures of Whites and negative words	40

IAT consisted of five classification tasks administered in a sequential order on the computer screen. Block 1, 2 and 4 were learning blocks, whereas blocks 3 and 5 were critical to compute IAT scores. The order of the critical blocls was counterbalanced across participants. In each block, participants were required to classify stimuli by pressing one of two different keys on the computer keyboard ('D' and 'K'). Instructions and key assignments were displayed on the computer screen before each block. Stimuli were presented at the center of the screen and remained visible until response.

IRI. The Italian version of the IRI is a self-report questionnaire composed of 28 items measuring affective and cognitive components of empathy. The questionnaire is organized in four 7-item subscales, two affective subscales and two cognitive subscales. The affective component is assessed by the empathic concern (EC) and personal distress (PD) subscales. The EC subscale taps the feelings of warmth, compassion, and concern for others. The PD subscale assesses the feelings of anxiety in situations of social relations. The cognitive component is assessed by the perspective-taking (PT) and fantasy subscales (FS). The PT subscale measures the tendency to spontaneously adopt the psychological point of view of others. The FS subscale measures the tendency to imaginatively transpose oneself into fictional situations (e.g., like movies or book characters). Both the PT and FS subscales have been shown to be positively correlated with other validated measures of cognitive empathy, such as the Hogan (1969) empathy scale.

EEG acquisition and analysis. EEG activity was recorded from 64 electrodes distributed over the scalp according to an extension of the international 10/20 system referenced to the left earlobe. The EEG was re-referenced offline to the average of the left and right earlobes. Trials contaminated by eye blinks, large horizontal eye movements or incorrect responses in the pain decision task were discarded from analysis.

Mean N1, P2, and N2 component amplitudes were measured at fronto-central electrode sites in 80–100 ms, 120–150 ms, and 200–240 ms time-windows locked to the onset of the face stimuli, respectively. The mean P3 amplitude was measured in a 400–750 ms time-window at Pz, P3 and P4 electrode sites. An additional 280–340 ms window, spanning the trough between N2 and N3 peaks, was selected based on visual localization of factor effects maximum values. Statistical analyses were conducted on individual amplitude estimates of activity recorded at each of frontal (AF3/AF4, AF7/AF8, Fz, FCz, F1/F2, F3/F4, F5/F6, F7/F8, FC1/FC2, FC3/FC4, FC5/FC6, FT7/FT8), central (Cz, C1/C2, C3/C4, C5/C6), and parietal (Pz, P3/P4) electrode sites. In all multi-factorial analyses, a Greenhouse-Geisser correction was used where appropriate.

The standardized Low Resolution Brain Electromagnetic Tomography (sLORETA) (Pascual-Marqui, 2002) was used for brain localization of the potential sources of ERP reactions to pain. sLORETA analyses were conducted following the creation of a boundary element method (BEM) model, with about 5000 nodes from MRI data, the selection of a temporal window in which ERP responses differentiated between painful and non-painful stimulations, and a location-wise inverse weighting from the minimum norm least square (MNLS) analysis with estimated variances.

3.2 Results

Pain decision task. Reaction times (RTs) exceeding each individual mean RT in a given condition +/-2.5 *SD* and RTs associated with incorrect responses were excluded from the RT analysis. Individual mean proportions of correct responses and RTs were submitted to separate repeated measure analyses of variance (ANOVAs), both considering stimulation condition (painful vs. non-painful) and race (White faces vs. Black faces) as within-subjects factors. Neither ANOVA showed significant factor effects (max $F = 2.2$; min $p = .16$).

IRI. Scores were computed by summing 1–7 scores to each item of the four subscales. Inter-individual mean rating scores were 27.1 ($SD = 4.03$) for the PT subscale, 25.0 ($SD = 4.94$) for the FS subscale, 26.8 ($SD = 2.96$) for the EC subscale, and 21.4 ($SD = 5.70$) for the PD subscale.

IAT. *Score* calculation followed the *improved algorithm* proposed by Greenwald et al. (2003). *D* positive scores were taken to indicate a preference towards White people. Mean *D* scores were significantly different from zero (mean *D*+/-standard error: 0.37+/-0.17; $t(11) = 2.2$, $p < .05$), reflecting a successful detection of a positive bias towards own-race members.

ERP: N1, P2, and N2. Figure 2 shows mean amplitudes of the subset of face-locked ERP components recorded at electrode site Fz that previous studies have indicated as sensitive to race, namely, N1, P2, and N2. We submitted to ANOVA individual amplitude values of each component considering stimulation (painful vs. non-painful), race (White faces vs. Black faces), and electrode site as within-subjects factors. The ANOVAs carried out on N1, P2 and N2 mean amplitude values revealed main effects of race over all electrodes of the frontal area (all *p*s < .05). As expected, Black (other-race) faces elicited N1 and P2 of greater amplitude than White (own-race) faces, and White faces elicited an N2 of greater amplitude than Black faces. The ANOVAs detected no other main effect or interaction (all *p*s > .05).

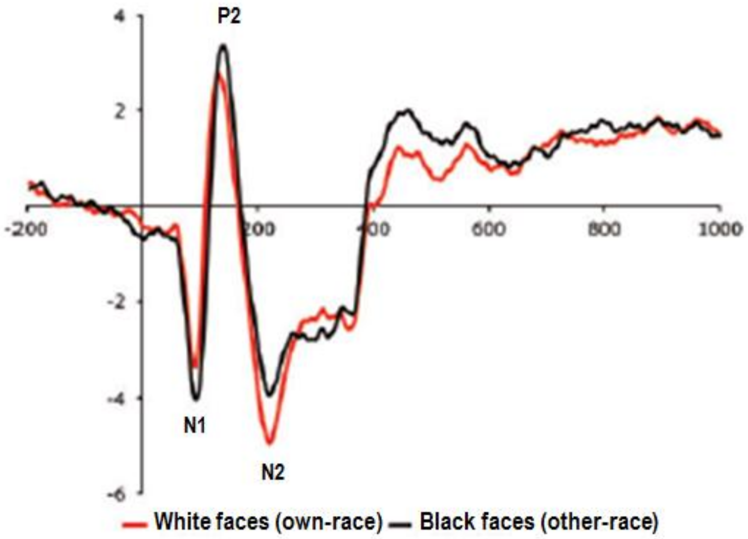


Figure 2. ERPs time-locked to the presentation of the faces recorded at electrode site Fz in response to white/own-race and Black/other-race faces collapsed across stimulation conditions (painful vs. non-painful).

ERP: N2–N3. Figures 3a and 4 show face-locked ERPs recorded from a selection of frontal electrode sites, separately for White and Black faces. A visual inspection of Figures 3a and 4 suggests that participants processed differently the faces in painful and non-painful conditions, exhibiting a positive shift in response to painful stimulation only when applied to White (own-race) faces relative to Black (other-race) faces. This observation was substantiated by an ANOVA carried out on individual ERP mean amplitude values recorded in a 280–340 ms time-window considering recording side (left vs. right hemisphere electrodes) as additional factor, which showed a significant interaction between stimulation condition and race on all the frontal and central electrode sites (min $F = 5.4$; max $p = .013$, min $\eta_p^2 = .516$). This pattern was bilaterally distributed, for recording site did not produce significant effects either as main factor or in interactions with race and stimulation conditions (all *F*s < 1; min *p* > .45). Planned comparison indicated that painful stimuli elicited a positive shift relative to non-painful stimuli in the N2–N3 time window only when applied to White (own-race) faces (min *t* = 2.39; max *p* = .036). This shift was absent when Black (other-race) faces were painfully stimulated (all *p*s < .23). These findings suggest that, in a temporal window of 280–340 ms, participants were in a state of experience sharing elicited by the presentation of suffering own-race members, but not of suffering other-race members. A scalp topographic map of N2–N3 activity elicited by White (own-race) faces in the painful stimulation condition is depicted in Figure 3b (upper panels).

The results of sLORETA analysis revealed that the neural activity in the N2–N3 time window that differentiated between painful and non-painful stimulations applied to White faces was mainly localized in the left IFG (Brodman area, BA, 45; peak MNI coordinates: -50, 25, 20, Fig. 3b, bottom), a core region of the mirror neuron system. This finding provides further support to the view that the early ERP reaction to pain was a likely reflection of neural resonance elicited by own-race faces in a painful condition.

We also correlated individual pain reactions detected in the N2–N3 temporal window for each electrode site over frontal and central areas with the ratings collected with the IRI subscales. Pain reactions in the present time-window were isolated by subtracting ERPs elicited in the non-painful stimulation condition from ERPs elicited in the painful condition, separately for White (own-race) faces and Black (other-race) faces. Pain reactions recorded from a subset of frontal electrodes positioned on the left hemisphere (i.e., AF7, AF3, F7, F5, F3, F1, FT7, FC5) correlated with the EC ratings (measuring the affective component of empathy), but only when White faces were presented. Spearman *r*s ranged from .50 at electrode site F7 to .73 at electrode site AF7 (all *p*s < .05). Figure 5a shows the scatterplot of individual EC ratings and pain effects recorded at electrode site AF7, and a graphical indication of the additional electrode sites where EC ratings and pain effects were significantly correlated. Pain reactions in this temporal window, however, did not correlate with IAT *D* score (all *p*s > .05). Figure 5b shows the scatterplot of individual EC ratings and pain reactions recorded at the electrode site AF7 for Black faces. At all frontal electrode sites, the correlations between EC ratings and pain reactions elicited by Black faces were not significant (*r*s ranged from .33 to .45 all *p*s > .05).

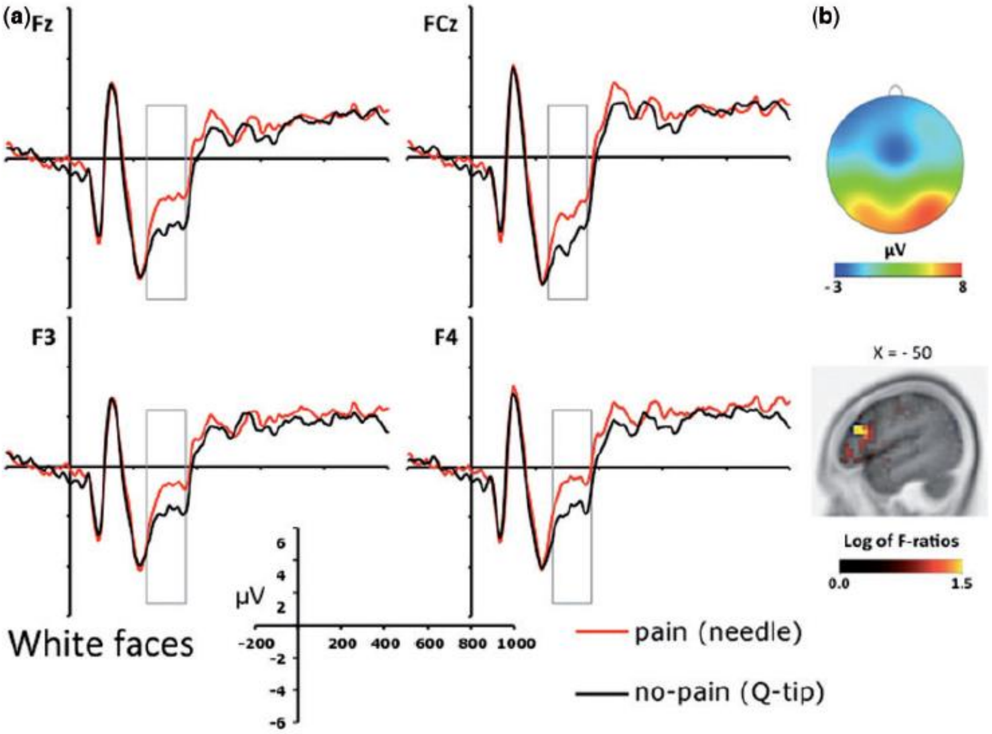


Figure 3. a) ERPs recorded at a selection of frontal electrode sites (i.e. Fz, FCz, F3 and F4) relative to the two stimulation conditions (painful vs. non-painful) for White/own-race faces. b) Voltage topography of the N2-N3 activity recorded in the painful condition (upper panel) and source estimation of the N2-N3 activity in the painful vs. non-painful conditions for White/own-race faces (lower panels).

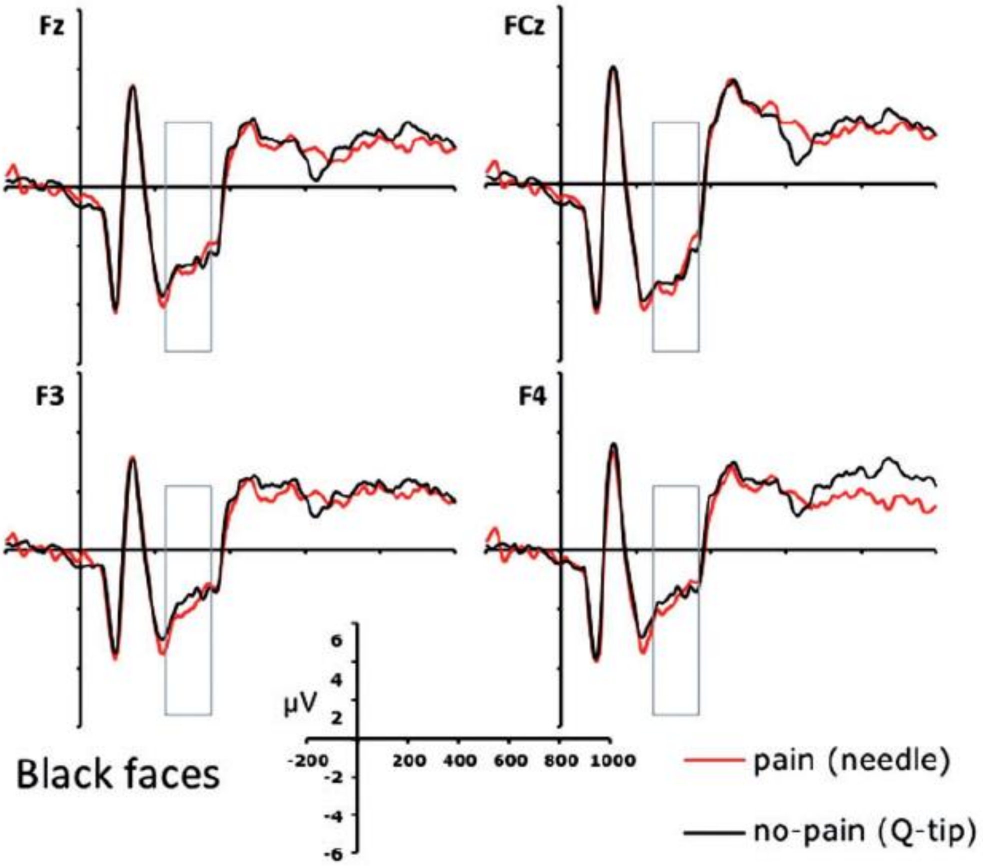


Figure 4. ERPs recorded at a selection of fronal electrode sites (i.e. Fz, FCz, F3 and F4) relative to the two stimulation conditions (painful vs. non-painful) for Black/other-race faces.

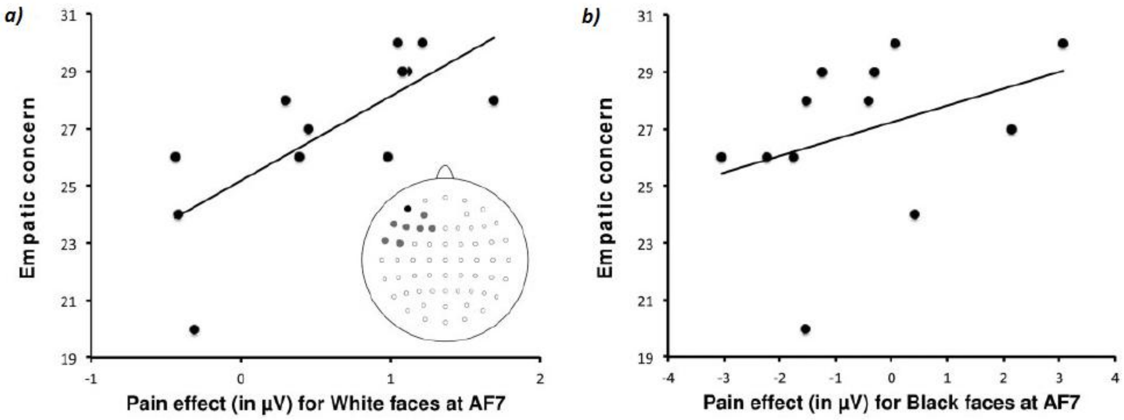


Figure 5. a) Scatterplot of the correlation between individual EC ratings and the pain effect observed at the electrode site AF7 in the time range of the N2-N3 components for White faces. Pain effect was computed by subtracting waveforms elicited in the non-painful stimulation condition from the waveforms elicited in the painful condition. A schematic illustration of the EEG montage shows the electrode sites on the scalp for which correlations were significant (AF7 in Black color, AF3, F7, F5 F3, F1, FT7 and FC5 in gray color). b) Scatterplot of the

correlation between individual EC ratings and the pain effect observed at the electrode site AF7 in the time range of the N2-N3 components for Black/other-race faces.

ERP: P3. An ANOVA on individual P3 amplitude values recorded at Pz electrode site revealed a main effect of race. P3 amplitude was greater for Black (other-race) faces (4.4 µV) than White (own-race) faces (3.3 µV; $F(1, 11) = 11.38$, $p < .01$, $\eta_p^2 = .532$). The main effect of stimulation condition was also significant ($F(1, 11) = 7.05$, $p < .05$, $\eta_p^2 = .414$). Notably, the interaction between these two variables did not reach significance ($F < 1$), suggesting no role of race in modulating a neural reflection of the cognitive component of empathy. P3 pain reactions were again isolated by subtracting amplitude values in the non-painful stimulation condition from amplitude values in the painful condition, separately for White (own-race) faces and Black (other-race) faces. P3 pain reactions amounted to .80 µV and .94 µV for White and Black faces, respectively. An analogous pattern was observed when separate ANOVAs were carried out on P3 mean amplitude values recorded at P3 and P4 electrode sites. Both ANOVAs indicated main effects of race (P3: $F(1, 11) = 7.45$, $p < .05$, $\eta_p^2 = .404$; P4: $F(1, 11) = 6.27$, $p < .05$, $\eta_p^2 = .363$), and marginally significant main effects of stimulation condition (P3: $F(1, 11) = 4.09$, $p = .068$, $\eta_p^2 = .271$; P4: $F(1, 11) = 3.58$, $p = .085$, $\eta_p^2 = .246$). No interaction between race and stimulation condition was observed ($F < 1$, $p > .7$). Face-locked P3 components recorded at Pz in painful and non-painful stimulation conditions are shown in Figure 6a for White (own-race) faces Black (other-race) faces.

Figure 6b shows scalp topographic maps of P3 activity elicited in the painful stimulation condition for White faces and Black faces and the source estimations of the P3 that differentiated between painful and non-painful stimulations applied to White and Black faces. The results of the sLORETA analysis suggested that reactions to the pain of White own-race individuals and Black other-race individuals in the P3 time window had potential sources in different brain regions underpinning the mentalizing system. The left middle frontal gyrus (MFG; BA 9; peak MNI coordinates: -30, 35, 40) was the likely source of the reaction to the pain of own-race individuals. The left temporo-parietal junction (TPJ; BA 40; peak MNI coordinates: -60, -50, 20; BA 22; peak MNI coordinates: -63, -40, 20) was the likely source of the reaction to the pain of other-race individuals. Since both these brain regions are involved in the attribution of mental states to others (e.g., Amodio & Frith, 2006; Saxe & Kanwisher, 2003; Frith & Frith, 1999; Moriguchi, Ohnishi et al., 2006; Spiers & Maguire, 2006), these results suggest that participants devolved the same amount of cognitive resources while mentalizing others' pain.

P3 reactions to pain did not correlate with any of the IRI subscales, including the PT and the FS subscales. Differential P3 pain reactions between White and Black faces recorded at Pz, P3 and P4 electrode sites did not correlate with IAT *D* scores ($p > .38$). Since the different potential sources of the P3 reactions to the pain of own-race and other-race individuals could suggest that these reactions are differently sensitive to the implicit pro-ingroup preference, we decided to compute correlations between P3 pain reactions and IAT *D* scores for White and Black faces, separately. Interestingly, P3 pain reactions recorded at P3 electrode site correlated with IAT *D* scores for White (own-race) faces ($r = .512$, $p < .05$), but not for Black faces ($p > .10$; see Figures 7a and 7b), likely reflecting a link between a pro-White (i.e., pro-ingroup) attitude and the magnitude of the P3 reaction to pain in the left hemisphere. No correlations were found between P3 pain reactions recorded at Pz and P4 electrode sites and IAT *D* scores ($p = .153$ and $p = .499$, respectively).

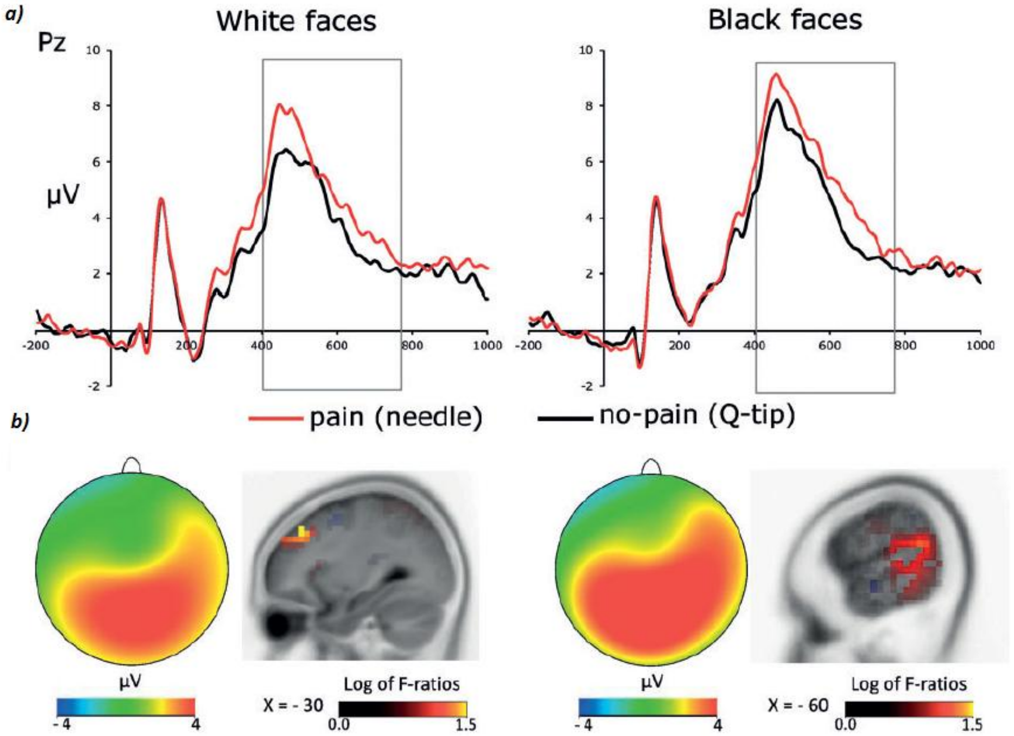


Figure 6. a) P3 ERP component recorded at the electrode site Pz relative to the two stimulation conditions (painful vs. non-painful) for White/own-race faces and for Black/other-race faces. b) Voltage topographies and source estimation of P3 activity as a function of race and stimulation conditions (painful vs. non-painful).

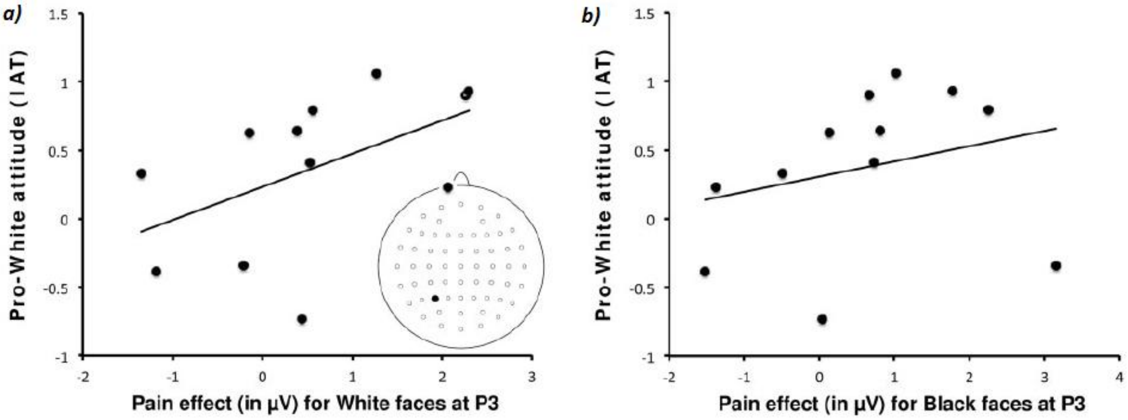


Figure 7. a) Scatterplot of the correlation between individual IAT scores and the pain effect observed at the electrode site P3 in the time range of the P3 components for White/own-race faces. Pain effect was computed by subtracting waveforms elicited in the non-painful stimulation condition from the waveforms elicited in the painful condition. b) Scatterplot of the correlation between individual IAT scores and the pain effect observed at the electrode site P3 in the time range of the P3 components for Black/other-race faces.

3.3 Discussion

The results of Experiment 2 showed that neural reactions to the pain of own-race individuals recorded in a 280–340 ms time-range, spanning the N2 and N3 ERP components, were magnified relative to neural reactions to the pain of other-race individuals. However, neural reactions of comparable magnitude towards the pain of own-race and other-race individuals were observed in a 400–750 ms time-window, a time-range typically associated with the P3 ERP component. In line with previous studies using similar experimental paradigms (Avenanti et al., 2010; Decety et al., 2010; Fan & Han, 2008; Li & Han, 2010; Xu et al., 2009), these neural reactions to pain have been interpreted as reflecting the engagement of distinct empathy-related subprocesses: Experience sharing and mentalizing (see also Experiment 1 described in Chapter 2).

Consistently with findings of Experiment 1 and as suggested previously by Fan and Han (2008), I argue that pre-P3 ERP reaction to pain can be ascribed to the experience sharing component of empathy, as further corroborated by the correlation between such ERP modulations and the EC ratings of the IRI. The set of correlations was particularly evident at left-frontal electrode sites, nicely dovetailing with recent fMRI work reporting analogous correlations between EC ratings and BOLD responses recorded from the left insula and ACC (Singer et al., 2004). The likely source of this ERP response was found in the left IFG, a key region of the mirror neuron system, thus dovetailing with prior similar evidence (e.g., Chakrabarti et al., 2006; Frühholz & Grandjean, 2012; Jabbi & Keysers, 2008; Minio-Paluello et al., 2006; Schulte-Ruther et al., 2007; Shamay-Tsoory et al., 2004, 2009).

In the P3 time-range, the racial bias on neural reactions to pain was reduced to nil. The source of P3 responses to pain was localized in the MFG and in the TPJ for own-race faces and other-race faces, respectively, in line with the cognitive functional characterization of this activity. The MFG is held to be involved in mentalizing (e.g., David et al., 2008; Lombardo, Chakrabarti, Bullmore, & Baron-Cohen, 2011; Moriguchi et al., 2006; Ohnishi et al., 2004; Spiers & Maguire, 2006). The TPJ is crucial in detecting goals from observed behavior, in self/others distinctions (Frith & Frith, 2001) and in representing beliefs of other individuals (e.g., Saxe & Wexler, 2005).

In their meta-analysis, Lamm, Decety and Singer (2011) pointed out the pervasive bilateral activation of AI and medial/anterior cingulate cortex whenever an empathic response is elicited. Additional brain structures are however engaged based on the stimuli used to elicit an empathic reaction. More specifically, while realistic empathy-eliciting stimuli (e.g., body parts in painful situations) engage selectively IFG and MFG (among other regions), abstract visual symbols (e.g., colors displayed on a screen indicating a painful stimulation) engage preferentially TPJ (among other regions). Lamm et al. (2011) proposed that these two different circuitries concur to the empathic state throughout two different pathways. The first, including IFG and MFG, allows action understanding that in turn triggers inferences about potential (affective) consequences of observed actions. The second pathway, including TPJ, subserves mentalizing and autobiographical memory, allowing the understanding of others' mental states on the basis of inferences based on one's own past experience. These considerations suggest that P3 reactions to the pain of own-race and other-race individuals are qualitatively different.

In the present study, P3 reactions to pain of White own-race faces correlated with IAT *D* scores although that was confined to the left hemisphere. This finding bears a close conceptual analogy with results by Avenanti et al. (2010), who showed reduced sensory-motor contagion following the exposure to other-race individuals' pain in participants exhibiting a higher degree of implicit pro-ingroup preference. Conversely, an analogous correlation was absent when we considered pre-P3 ERP reactions to pain.

This finding suggests that, in the P3 time-window, participants' implicit pro-White preference interacted with the process of evaluation of the painful condition, such that the evaluation tended to be greater for participants who showed a particularly high implicit pro-White preference. However, the interactive effect did not spread throughout the scalp, and the P3 recorded at the other electrode sites (Pz and P4) did not correlate with the IAT *D* scores, suggesting that the overall evaluation of painful stimuli was largely independent on the implicit pro-White preference.

The present results complement and extend previous fMRI (Xu et al., 2009) and TMS (Avenanti et al., 2010) work by mapping out the time-course of the temporally asynchronous engagement of an early neural resonance component of empathy, which amplifies responses to the pain of own-race members, and of a cognitive component likely related to mentalizing which magnitude appears not to be influenced by their race, although the underlying different potential source estimates suggest that this later reaction qualitatively differs between own-race and other-race conditions.

Chapter 4

Experiments 3: Trustworthiness implicit appraisal

4.1 Experiment 3: Visual working memory for trustworthy and untrustworthy faces

In Experiment 3, I adopted a different paradigm, namely the change detection task, and the SPCN ERP component, to test whether trustworthiness may modulate visual working memory (i.e. VWM) processing under condition of task-irrelevance. That was necessary prior to investigate whether implicit perceived trustworthiness affects neural empathic responses (i.e. aim of the Experiment 4 and 5).

Due to the specificity of the object of the current experiment, i.e. the maintenance of implicitly perceived trustworthiness' faces in VWM, I'll go into more detail in the introduction and in the discussion section of the present chapter.

As I described in Chapter 1, trustworthiness is one of the first evaluation people constantly perform when facing strangers at "first sight" (e.g., Adolphs, 2002; Baron, Gobbini, Engell, & Todorov, 2011; Engell, Haxby, & Todorov, 2007; Oosterhof & Todorov, 2008; Stirrat & Perrett, 2010; Todorov, 2008; Todorov, Baron, & Oosterhof, 2008; Willis & Todorov, 2006). This is not surprising, especially from an evolutionary perspective, since assessing faces' trustworthiness is a critical social tool in order to avoid untrustworthy individuals (and consequent risky social interactions), and to approach trustworthy individuals for immediate or future cooperation (e.g., Boone and Buck, 2003; Oosterhof and Todorov, 2008; Trivers 1971; Zebrowitz and Montepare, 2005). This evolutionary view is supported by empirical evidence linking trustworthiness and approachability appraisals (Todorov, 2008; Santos and Young, 2008a, 2008b).

As the reader might remember, a conspicuous body of behavioral research examined the physical characteristics that guide people in trustworthiness evaluation (e.g., Knutson, 1996; Montepare & Dobish, 2003; Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008). To this end, Oosterhof and Todorov (2008) developed a data-driven statistical model individuating facial features related to judgments of trustworthiness: High inner eyebrows, pronounced cheekbones, wide chins and shallow nose sellion, characterize faces appearing trustworthy while faces evaluated as untrustworthy are characterized by low inner eyebrows, shallow cheekbones and thin chins and deep nose sellion. While this behavioral model provides a clear indication of which facial cues are involved in trustworthiness appraisal, researchers have more recently begun to investigate the neural underpinnings of such facial evaluation. Thus far these investigations have primarily utilized functional magnetic resonance imaging (fMRI; e.g., Adolphs, Tranel, & Damasio, 1998; Adolphs, 2002; Engell, Haxby, & Todorov, 2007; Said, Baron, & Todorov, 2008; Winston, Strange, O'Doherty, & Dolan, 2002). Two recent meta-analyses (Bzdok, Langner, Caspers, Kurth, Habel, Zilles, Laird, & Eickhoff, 2011; Mende-Siedlecki, Said, & Todorov, 2012) summarized these findings as revealing differential roles of brain circuitries in processing trustworthy and untrustworthy faces, with the former class of faces engaging reward-associated brain regions (including the nucleus accumbens) and untrustworthy faces engaging a brain region responding to potential threat, i.e. the ventral portion of amygdala. In general, these favor a key role of approach and avoidance motivation systems in reacting to faces characterized by different levels of trustworthiness as mentioned above (Chen and Bargh, 1999; Cosmides and Tooby 2000; Said et al. 2008; Todorov, 2008). In line with this evidence, people may notice, remember, or interpret the same social information quite differently, in particular when facing *ambiguous* social cues. This observation suggests that individuals may differ in their perception of a stranger as innocuous or, rather, as a potential offender. Notably, compelling evidence suggests that anxiety may play a crucial role in this context since individual differences in anxiety levels are associated with differences in the distribution of cognitive resources (e.g., Hirsch & Clark, 2004; Holmes, Nielsen, Tipper & Green, 2009; Moser, Huppert, Duval & Simons, 2008; Rossignol, Philippot, Bissot, Rigoulut & Campanella, 2012) such that high-anxious individuals tend to allocate excess attention and working memory resources to threat-related cues and to misinterpret emotionally ambiguous stimuli as more negative compared to non-anxious individuals (see Mathews & MacLeod, 1994; 2005 for reviews; Calvo, Eysenck, & Castillo, 1997; Klumpp, Angstadt, Nathan, & Phan, 2010; Mathews & Mackintosh, 1998; Stout, Shackman, & Larson, 2013; Yoon & Zinbarg, 2007). Incidentally, high levels of anxiety are closely linked to avoidance motivation (e.g., Gable, Reis, & Downey, 2003).

On the basis of these findings, the main hypothesis of this study was that implicit perceived trustworthiness may modulate visual working memory (i.e., VWM) faces' representations and, as an important corollary to this, that individual differences in anxiety (either general and social anxiety) would further modulate the resolution(i.e., precision) of these representations.

To investigate these questions I monitored ERPs while participants performed a change detection task that required encoding and maintaining for a short interval the identity of standardized either trustworthy and untrustworthy faces. The SPCN, time-locked to the onset of the face, was used to index the resolution of face representations in VWM.

Blood-oxygen-level-dependent (i.e., BOLD) fMRI signal is characterized by a particularly low temporal resolution and is not suitable to investigate the time-course of trustworthiness appraisal. Event-related brain potentials (ERPs) provide instead high-resolution measures of the time-course of neural activity patterns associated with perceptual and cognitive processes. In the context of trustworthiness appraisal, only a few recent ERP studies have explored which stages of processing are sensitive to physical cues of faces' trustworthiness (Dzhelyova, Perrett, & Jentzsch, 2012; Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2012; Rudoy & Paller, 2009; Yang, Qi, Ding, & Song, 2011). Modulations of face processing driven by explicit trustworthiness appraisal occur as early as the C1 and P1 components time-locked to face onset (Marzi et al., 2012; Yang et al., 2011). Faces' trustworthiness continues to modulate cascading neural activity during early selection of visual stimuli with affective and motivational significance (as reflected in early posterior negativity modulations, i.e. EPN; Dzhelyova et al., 2012; Marzi et al., 2012), later structural encoding (as reflected in N170 modulations; Dzhelyova et al., 2012), and higher-order stages of processing as reflected in modulations of a fronto-central positivity (Marzi et al., 2012; Rudoy & Paller, 2009) and late positive component (i.e., LPC; Yang et al., 2011; Marzi et al., 2012). However, these studies did not illuminate whether the trustworthiness of a face may be implicitly appraised when individuals are simply exposed to such stimuli and, at the present, it remains unknown whether exposure to faces characterized by different levels of trustworthiness/untrustworthiness might reflect individual differences in individual anxiety.

As mentioned before, the present investigation focused on a privileged stage of face processing, i.e. VWM, since it acts as a cognitive hub (Haberlandt, 1997) for low-level processes − by which physical cues of trustworthiness are first encoded − and higher-order cognitive processes including decision-making, and long-term memory (Luck, 2008). In particular, I aimed to elucidate whether the amount of VWM resources were differently allocated to untrustworthy vs. trustworthy faces and whether that might be also subjective to individual differences in anxiety level. I adopted faces included in the database created according to the method described by Oosterhof and Todorov (2008; see also Todorov & Oosterhof, 2011). Such faces vary along the trustworthiness dimension +/- 2 SD and +/- 3 SD from neutral. Incidentally, the vast majority of strangers an observer continuously comes across is more likely perceived as moderately trustworthy or untrustworthy, therefore, understanding whether these moderately trustworthy/untrustworthy faces are differently represented in VWM has a relevant ecological significance.

The VWM task used in the current study was a modified version of the change detection task (e.g., Vogel & Machizawa, 2004; Sessa, Luria, Gotler, Jolicoeur, & Dell’Acqua, 2011; Sessa, Tomelleri, Luria, Castelli, Reynolds, & Dell'Acqua, 2012) and required participants to memorize face identities without an explicit trustworthiness evaluation, emphasizing the ecological validity of the task. I monitored the SPCN[15] (also labeled contralateral delay activity, CDA; Vogel & Machizawa, 2004) time-locked to faces recorded at posterior parietal sites indexing VWM maintenance, component of the ERP. Sessa et al. (2011, 2012) demonstrated that SPCN amplitude varies proportionally to the resolution of faces’ representations in VWM, such that high-

---

[15] The SPCN is computed as the difference between contralateral activity and ipsilateral activity time-locked to a lateralized target stimulus. Its amplitude correlates positively with VWM informational load (e.g., Jolicœur, Brisson, and Robitaille, 2008; Perron, Lefebvre, Robitaille, Brisson, Gosselin, Arguin and Jolicœur, 2009; Robitaille, Grimault and Jolicœur, 2009; Vogel and Machizawa, 2004) and it has been shown to increase as the number (Vogel and Machizawa, 2004), and the complexity (Luria, Sessa, Gotler, Jolicœur, and Dell’Acqua, 2010) of stimuli to be held in VWM is increased up to the level of VWM saturation at which point the SPCN component usually tends to reach an asymptote.

resolution faces' representations elicit larger SPCN amplitudes relative to low-resolution faces' representations.

I predicted that trustworthy and untrustworthy faces would have been represented in VWM with a different resolution and that this difference (computed by subtracting SPCN elicited by trustworthy and untrustworthy faces) would have been related to individual differences in personality traits, i.e. the level of participants' anxiety, such that higher anxious participants would have maintained higher-resolution representation of untrustworthy faces as compared to lower anxious participants. To this aim, at the end of the ERP recording session, participants were also administered the State Trait Anxiety Inventory (Spielberger, 1983) and the Italian version of the Social Interaction Anxiety Scale and the Social Phobia Scale questionnaires (Sica et al., 2007; for the English version refer to Mattick & Clarke, 1998).

4.1.1 Method

Participants

Data were collected from 16 healthy volunteer students (3 males) from the University of Padova (mean age: 24.56 years, SD = 1.63) who reported normal or corrected-to-normal vision and no history of neurological disorders. All participants gave their informed consent according to the ethical principles approved by the University of Padova. Data from 4 participants (all females) were discarded from the analyses because of an excessive rate (higher than 30% of trials) of EEG artifacts.

Stimuli and procedure

Memory task. Prior to the ERP experiment an independent student sample (N = 30; 12 males, mean age: 23.43 years, SD = 1.86) provided 7-step ratings of facial trustworthiness (trustworthy versus untrustworthy) and emotional expression (happy versus angry) of 110 neutral facial expression identities generated using FaceGen Modeller 3.2 (Singular Inversions, 2007; +/-2 and +/-3 SD from neutral) according to the methods described by Oosterhof and Todorov (2008). This procedure allowed selecting the most appropriate face stimuli for the present investigation resulting in 10 untrustworthy (-2 and -3 SD) and 10 trustworthy (+2 and +3 SD) bald Caucasian male faces with a non-significant correlation with the emotion scale (consensus neutral expression $p > .1$; for a similar procedure see Yang et al., 2011).

The face stimuli were scaled using image-processing software so that each face fitted in 3.3° x 4.5° (width x height) rectangle from a viewing distance of approximately 70 cm. Face stimuli were randomly selected and memory-display and test-display were composed of 2 faces − with either trustworthy or untrustworthy facial characteristics − horizontally aligned and located at the upper or the lower two quadrants of an imaginary rectangle centered around fixation. The distance between the center of the face and the fixation cross was 4.9°.

Examples of two different trials are reported in Figure 1a and b and described in detail in the respective captions. The memory-display consisted of two faces displayed in each visual hemifield, preceded by arrow cues pointing to the side of the to-be-memorized face. The face located in the opposite hemifield had to be ignored. Following the memory-display, participants were required to examine the same pre-cued side of the test-display for a possible change in the identity of the face, which occurred unpredictably on 50% of trials. When a change occurred, the face was replaced with a different face of the same level of trustworthiness. The experiment consisted of 192 trials per condition (trustworthy vs. untrustworthy; 8 blocks of 48 trials each).

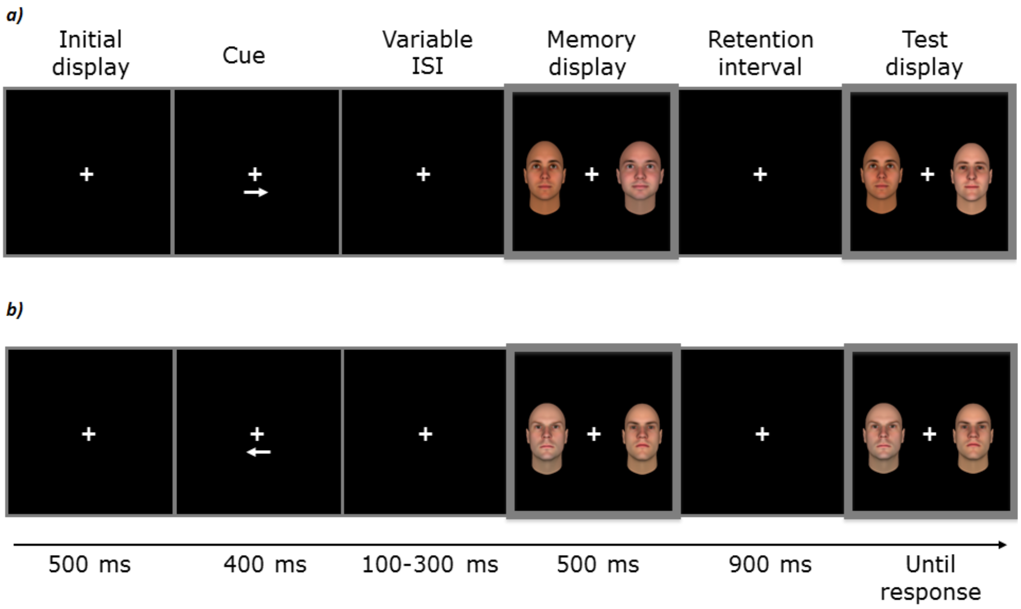


Figure 1. Examples of change detection task trials when either a) a trustworthy face (example for the right hemifield) or b) an untrustworthy face (example for the left hemifield) had to be encoded. ISI: interstimulus interval.

At the end of the ERP recording session, participants completed in hardcopy the State Trait Anxiety Inventory (STAI; Spielberger, 1983) and the Italian version of the Social Interaction Anxiety Scale (SIAS) and the Social Phobia Scale (SPS) questionnaires (Sica, Musoni, Chiri, Bisi, Lolli & Sighinolfi, 2007; for the English version see Mattick, & Clarke, 1998).

STAI. The Italian version of the STAI is a 20-item self-report questionnaire (for each form: Y-1, state and Y-2, trait) that measures a temporary (state: e.g. “I feel at ease”, “I feel upset”) or persisting (trait: e.g. “I am a steady person”, “I lack self-confidence”) emotional state of generalized anxiety. Participants responded using a 4-point Likert-type scale.

SIAS and SPS. The Italian version of the SIAS is a 19-item self-report questionnaire that measures the fear of social interaction situations (e.g. “When mixing socially, I am uncomfortable”, “I am nervous mixing with people I don’t know well”). The Italian version of the SPS is a 20-item self-report questionnaire that measures the fear of being evaluated or observed by unknown people during daily activities (e.g. “I worry about shaking or trembling when I'm watched by other people”, “I am worried people will think my behavior is odd”). Participants responded using a 5-point Likert-type scale.

Electrophysiological recording and analysis

EEG was recorded during the change detection task from 32 active electrodes distributed over the scalp in accordance with the international 10/20 system, placed on an elastic Acti-Cap referenced to the left earlobe. The EEG timeseries was re-referenced offline to the average of the left and right earlobes. Trials contaminated by eye blinks, large horizontal eye movements or incorrect responses in the change detection task were discarded from analysis. We computed contralateral waveforms by averaging the activity recorded at right hemisphere electrodes when participants were cued to encode the face stimulus on the left side of the memory-display with the activity recorded from the left hemisphere electrodes when they were cued to encode the face stimulus on the right side of the memory-display. SPCN was quantified at posterior electrodes sites (P7/P8) as the difference in mean amplitude between the ipsilateral and contralateral waveforms in a time window of 500–1100 ms relative to the onset of the memory array (i.e., the SPCN mean amplitude was quantified in a time-window following the disappearance of the memory-display).

We computed, for each participant and condition, an *SPCN trustworthiness score* reflecting differential resolution of VWM representations of trustworthy and untrustworthy faces based on the following equation:

$SPCN\ trustworthiness\ score = SPCN_{Trustworthy\ faces} - SPCN_{Untrustworthy\ faces}$

SPCN is a negative-going ERP response. Positive *SPCN trustworthiness scores* (SPCN differences) thus indicated that more information was encoded in VWM from untrustworthy faces compared to trustworthy faces (i.e., VWM advantage for representations of untrustworthy faces) and negative *SPCN trustworthiness scores* indicated the opposite (i.e., VWM advantage for representations of trustworthy faces).

4.1.2 Results

Memory task

VWM performance was quantified using a standard index of sensitivity (*d'*; Green & Swets, 1974). This measure allowed estimating how sensitive the participants were to changes between the memory and test-displays and whether this sensitivity differed as a function of faces' trustworthiness. These values were submitted to paired-sample t test considering the independent variable face trustworthiness (trustworthy faces, i.e. including faces +2 and +3 SD from neutral, vs. untrustworthy faces, i.e., including faces -2 and -3 SD from neutral). Analysis revealed that participants were equally accurate in responding to trustworthy (*mean d'* = 2.43, *SD* = .57) and untrustworthy (*mean d'* = 2.38, *SD* = .64) faces, $t < 1$. To better characterize these findings, we performed a backward stepwise regression analysis in which variables were sequentially removed from a full model (including *STAI Y-1*, *STAI Y-2*, *SPS* and *SIAS* as predictors, and *d'* values, for trustworthy and untrustworthy faces separately, as dependent variables). When *d'* for trustworthy faces was predicted, it was found that *STAI Y-1* ($\beta = .711$, $p < .005$) and *SPS* ($\beta = -.711$, $p < .005$) were significant predictors (adjusted $R^2 = .654$). For untrustworthy faces, SPS was a marginally significant predictor ($\beta = -.503$, $p = .075$;

adjusted $R^2 = .211$). In general, these findings suggest a pervasive effect of high levels of social phobia in deteriorating behavioral performance.

STAI. The mean rating scores was 37.13 (*SD* = 11.94) for state anxiety (form Y-1) and 41.19 (*SD* = 8.45) for trait anxiety (form Y-2).

SIAS and *SPS*. The mean rating scores was 17.06 (*SD* = 7.96) for SIAS and 13.44 (*SD* = 9.27) for SPS.

SPCN

Figure 2a illustrates contralateral and ipsilateral waveforms recorded at electrode sites P7/P8 time-locked to the memory-display for trustworthy and untrustworthy faces, separately. Figure 2b shows SPCN (contralateral minus ipsilateral) waveforms. An informal observation of waveforms indicates that SPCN was modulated by faces' trustworthiness.

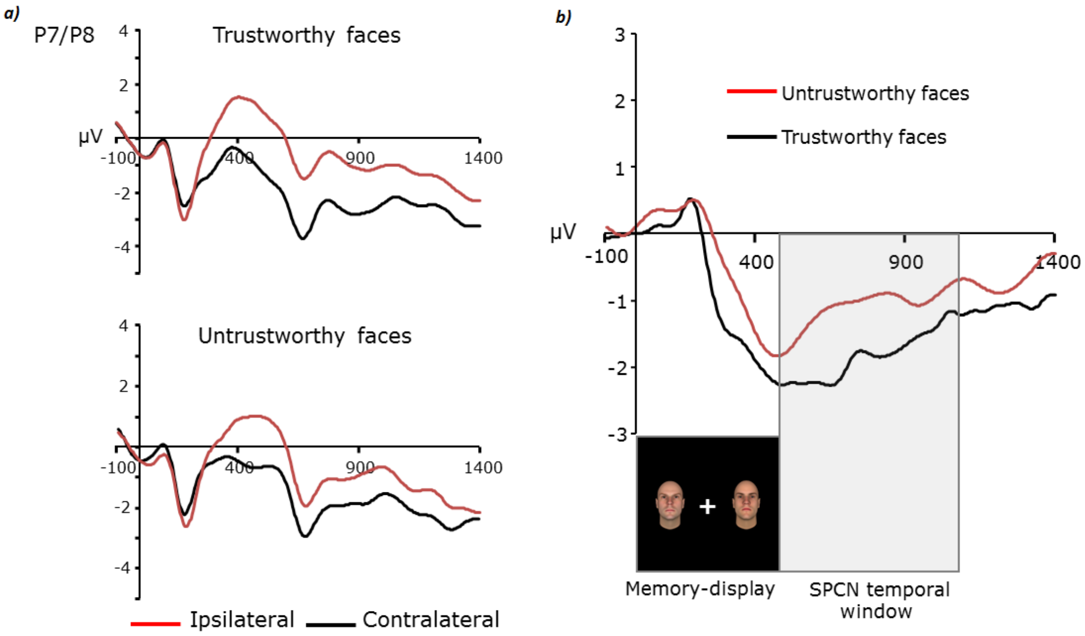


Figure 2. a) Contralateral and ipsilateral waveforms recorded at P7/P8 electrode sites time-locked to the onset of the memory-display for trustworthy and untrustworthy faces separately.

b) SPCN (contralateral-minus-ipsilateral) for trustworthy and untrustworthy faces, separately. The shaded area indicates the SPCN (contralateral-minus-ipsilateral waveforms) temporal window selected for statistical analyses (500–1100 ms).

SPCN mean amplitude values recorded at electrode sites P7/P8 were submitted to a paired-sample t test considering the independent variable face trustworthiness (trustworthy vs. untrustworthy). Trustworthy faces elicited larger SPCN amplitude (-1.83 µV; *SD* = 1.05) than untrustworthy faces (-1.05 µV; *SD* = 1.36), $t(11) = 2.984$, $p < .012$, $\eta_p^2 = .447$, indicating that implicit evaluation of trustworthiness modulated VWM processing; in particular, under conditions of exposure to faces in the middle of the trustworthiness dimension, trustworthy faces were overall maintained as higher-resolution representations compared to untrustworthy faces. For completeness, we also analyzed mean SPCN amplitude values recorded at electrode sites P3/P4 and O1/O2. The pattern for P3/P4 was similar to that observed for P7/P8 (i.e., a larger SPCN amplitude for trustworthy faces than untrustworthy faces, $t(11) = 2.386$, $p < .036$, $\eta_p^2 = .341$). The trend was analogous for SPCN mean amplitudes at electrode sites O1/O2, but statistically not significant (t <1).

Notably, *SPCN trustworthiness scores* were highly correlated with the level of both participants' state anxiety (i.e., *STAI Y-1 scores*), $r = .812$, $p < .005$, and social anxiety (i.e., *SIAS scores*), $r = .631$, $p < .05$ (see Figure 3a and Figure 3b).

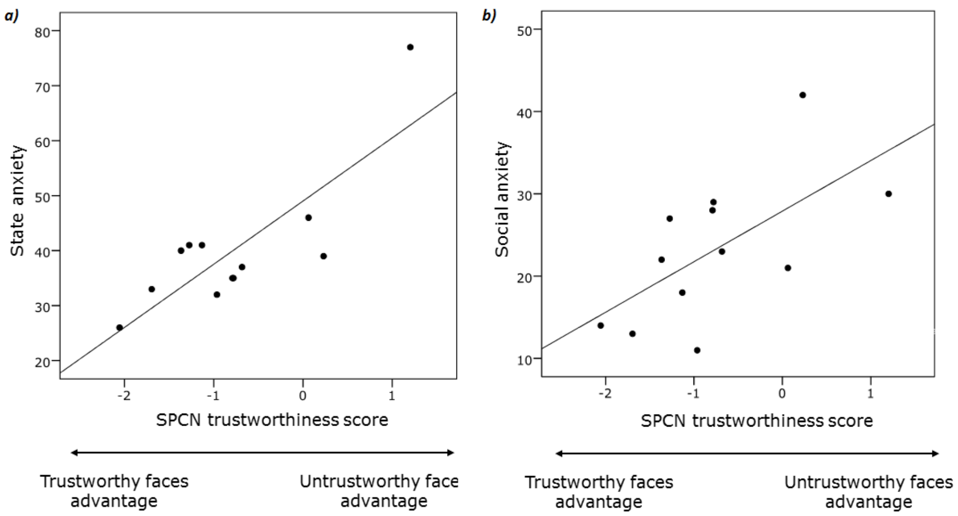


Figure 3. Correlations between SPCN trustworthiness scores and the level of participants’ (a) state anxiety ($r = .812$, $p < .005$) and (b) social anxiety ($r = .631$, $p < .05$).

The bar graph depicted in Figure 4 is presented only for illustrative purposes and illustrates mean SPCN amplitudes for trustworthy (blue bars) and untrustworthy (red bars) faces in participants with high and low levels of state and social anxiety, separately. The graph noticeably shows that trustworthy faces elicited similar SPCN amplitudes irrespective of the level of participants’ anxiety; on the other hand untrustworthy faces elicited increased SPCN amplitudes in high-anxious participants compared to low-anxious individuals.

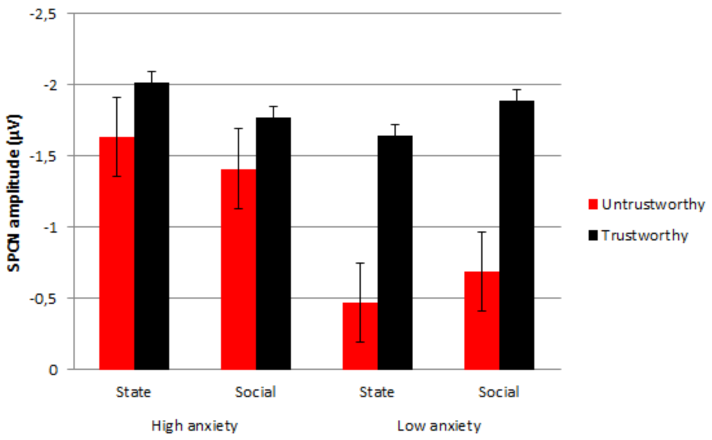


Figure 4. Bar graph of mean SPCN amplitudes for trustworthy (blue bars) and untrustworthy (red bars) faces for high and low levels of state and social anxiety, separately.

Based on the known properties of SPCN, these findings support the conclusion that the higher the anxiety self-reported by participants (in particular state and social anxiety) the higher was the resolution of VWM representations of untrustworthy faces (compared to lower-anxious participants). Finally, since the unbalance between the number of male (i.e., three) and female (i.e., nine) participants in this study might prevent interpreting our results as indicative of the general population, we analysed separately the effect of trustworthiness on the SPCN amplitude in male and female participants. Although not significant, both males and females showed a comparable tendency in showing increased SPCN amplitudes when trustworthy faces (vs. untrustworthy faces) had to be encoded (males: $t = -3.541$ $p = .072$; females: $t = -1.912$ $p = .092$), suggesting that, at least under the present experimental conditions, both males and females encoded trustworthy faces as higher resolution representations compared to untrustworthy faces.

4.1.3 Discussion

The results of the present experiment were clear-cut: The SPCN revealed that perceived trustworthiness is automatically maintained in VWM even when totally task-irrelevant. Further, individuals' level of general/social anxiety is related to the resolution with which faces characterized by features of trustworthiness and untrustworthiness (Oosterhof & Todorov, 2008) are represented in VWM.

The SPCN was monitored during maintenance of representations of parametrically manipulated trustworthy and untrustworthy faces that differed slightly from faces that were neutral on the trustworthiness dimension (Oosterhof & Todorov, 2008). Importantly, the memory task did not require explicit evaluation of faces' trustworthiness. On the basis of known properties of the SPCN, modulations of SPCN amplitude as a function of faces' trustworthiness denote differences in the resolution of those faces representations in VWM such that larger SPCN amplitudes reveal maintenance of higher-resolution representations (Sessa et al., 2011; 2012).

SPCN amplitude, on average, was increased for trustworthy faces compared to untrustworthy faces. This finding suggests that participants maintained in VWM trustworthy faces in higher-resolution representations as compared to untrustworthy faces. Notably, individual estimates of the difference in SPCN amplitude elicited by trustworthy vs. untrustworthy faces (i.e., *SPCN trustworthiness scores*) strongly correlated with individual estimates of state and social anxiety (i.e., *STAI Y-1 scores* and *SIAS scores,* $r = .812$ and .631, respectively), indicating that untrustworthy faces elicited larger SPCN amplitudes in high-anxious individuals than in low-anxious individuals. These findings provide novel evidence that VWM is sensitive to physical cues of trustworthiness and that this sensitivity is further modulated by individuals' anxiety.

We also observed overall larger SPCN amplitude for trustworthy faces relative to untrustworthy faces. As noted above, it has been suggested that trustworthiness appraisal, in particular for those faces *at the extremes* of the trustworthiness dimension, approximates the detection of emotional facial expressions (e.g., Oosterhof & Todorov, 2008). Congruent with these observations, ERP modulations for *extreme* trustworthy and untrustworthy faces nearly mimics ERP modulations for the corresponding facial emotional expressions (e.g., Marzi et al., 2012), indicating increased processing for highly untrustworthy faces compared to highly trustworthy faces.

To my knowledge, only two recent studies reported findings of memory tasks for trustworthy and untrustworthy faces (Rule, Slepian & Ambady, 2012; Todorov, Said, Oosterhof & Engell, 2011), providing contrasting evidence on which class of faces (trustworthy vs. untrustworthy) benefit from prioritized memory processing. Rule et al. (2012) have found a long-term memory advantage in terms of behavioral accuracy for untrustworthy faces compared to trustworthy faces. On the contrary, Todorov et al. (2011) reported a higher hit rate for trustworthy than untrustworthy faces in the context of a one-back recognition task. Notably, face stimuli used in those two investigations differed such that Rule et al. (2012) selected faces composing the trustworthy and untrustworthy face sets from faces that obtained highest and lowest trustworthiness scores on participants' ratings along a 7-point scale; conversely, and similarly to the present investigation, Todorov et al. (2011) used face stimuli in the middle of the trustworthiness dimension (i.e., +/- 1 SD and +/- 3 SD) generated on the basis of the approach of Oosterhof and Todorov (2008). The findings of the present investigation, in demonstrating a VWM advantage for moderately trustworthy faces relative to moderately untrustworthy faces, appear to complement previous evidence of prioritized processing of extreme untrustworthy faces (compared to trustworthy faces).

Indeed, congruent ERP work in this field also indicates that individuals with high levels of social anxiety (including both clinical and non-clinical anxiety) tend to process incoming emotional stimuli, in particular negative facial expressions, in greater detail compared to control groups at diverse stages of processing, as indexed by enhanced amplitude of P1 (e.g., Kolassa, Kolassa, Bergmann, Lauche, Dilger, Miltner, & Musial, 2009; Kolassa & Miltner, 2006), P2 (Rossignol, et al., 2012) and N170 components

(Kolassa & Miltner, 2006). In this vein, a recent ERP study using a similar task to that implemented in our study, monitored the SPCN component and demonstrated that anxiety is associated with inefficient gating of threat-related faces (i.e., fearful faces) from VWM even when task-irrelevant (Stout et al., 2013). Furthermore, even when stimuli are not characterized by a clear negative valence, high-anxious individuals tend to misinterpret these emotionally ambiguous stimuli as more negative compared to non-anxious individuals (Calvo & Castillo, 2001; Klumpp, Angstadt, Nathan, & Phan, 2010; Mathews & Mackintosh, 1998; Yoon & Zinbarg, 2007). To note, the memory task implemented in the present investigation required participants to encode faces' identities that did not express an emotion, although it has been proposed that trustworthiness appraisal approximates the detection of emotional facial expressions (Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz & Montepare, 2008), in particular for those faces *at the extremes* of the trustworthiness dimension, such that highly trustworthy and untrustworthy faces are perceived as happy and angry, respectively (Todorov, 2008).

These observations suggest that high anxiety individuals tended to perceive *moderately* untrustworthy faces as more untrustworthy compared to low anxiety individuals, leading to increased resolution of those VWM representations. This interpretation is consistent with previous evidence showing increased resolution of VWM representations of face identities with negative expressions (e.g., Sessa et al., 2011). The present findings suggest that the improved VWM processing of untrustworthy faces in high-anxious individuals (compared to low-anxious individuals) is related to a general oversensitivity towards potential threat regardless of the nature (either social or not) of the threat.

Taken together, these findings seem to reveal high flexibility of aversive/avoidance and appetitive/approach motivational systems (Chen & Bargh, 1999; Cosmides &

Tooby, 2000; Todorov, 2008) in reacting to social stimuli and consequently biasing downstream cognition, such as memory. This flexibility may rely on the notion of value-prediction code linked to a stimulus (e.g., Anderson, Laurent, & Yantis, 2011a, 2011b; Raymond & O'Brien, 2009) that combines information of both intensity/magnitude and valence of that stimulus and associated potential outcome (i.e., gain or loss, that in the context of trustworthiness appraisal translates into opportunities of cooperation vs. risky interactions). Within this theoretical framework it appears reasonable to hypothesize that when exposed to very untrustworthy faces, which also may appear as expressing anger, the high intensity and negative valence conveyed by those faces increases the likelihood of *reaching a threshold of threat detection such that* the value-prediction code assigned by individuals is weighted more highly than value assigned to trustworthy faces. This will then bias processing (including VWM maintenance) in favor of potential threat in the environment. Incidentally, this explanation fits nicely with previous behavioral and ERP work showing a VWM advantage for negative facial expressions (i.e., angry and fearful faces) compared to neutral and/or positive facial expressions (i.e., happy faces; Jackson, Wolf, Johnston, Raymond, & Linden, 2008; Jackson, Wu, Linden, & Raymond, 2009; Sessa et al., 2011). On the other hand, when perceived threat from untrustworthy faces is low, people may tend to assign a low value to them, and to assign a higher value and to allocate a larger proportion of cognitive resources to/on stimuli with rewarding characteristics (e.g., Raymond & O'Brien, 2009), such as trustworthy faces (e.g., Todorov, 2008), according to an approach behavior. Along this theoretical perspective, the findings of this study also advocate that value-prediction codes assigned to untrustworthy faces strongly depend on the levels of individuals' anxiety, such that higher levels of anxiety are associated with higher-resolution representations of untrustworthy faces (compared to low-anxious individuals).

Modulations of the SPCN as a function of trustworthiness were not accompanied by analogous behavioral effects. This dissociation between SPCN modulations and behavior is not alarming for at least two classes of considerations. First, SPCN is a pure measure of VWM representation; on the contrary the overt response required in the change detection task reflects not only the (quality) of current VWM representation, but also additional processes allowing to compare the face presented in the test-display with current VWM representation (see, for instance, Awh, Barton, & Vogel, 2007). This interpretation is also in line with the finding of deteriorating effect of social phobia selectively on behavioral performance, for both trustworthy and untrustworthy faces. Convincing evidence suggests that anxiety may deteriorate behavioral performance, in both matching tasks (i.e., Attwood, Penton-Voak, Burton, & Munafò, 2013) and VWM tasks (e.g., Hirsch & Mathews, 2008; Moriya & Sugiura, 2012). .

A second type of considerations refers to the observation that brain responses are often more sensitive to subtle processing differences than behavioral measures (e.g., Heil, Rolke, & Pecchinenda, 2004; Luck, Vogel, & Shapiro, 1996; Wilkinson & Halligan, 2004) and this dissociation may be particularly evident when the behavioral task requires a dichotomous response such that was required in the change detection task implemented in this study (Sessa et al., 2011, 2012).

To conclude, present findings provide evidence that physical cues of faces' trustworthiness modulate the quality of faces' representations in VWM even under conditions in which trustworthiness is implicitly appraised, and given the privileged position of VWM within the stream of processing, this finding may be particularly relevant for models of approach/avoidance motivational systems and behavior. Moreover, high levels of individuals' anxiety modulate VWM maintenance of those faces' representations.

On the basis of the results of this experiment, I adopted standardized Oosterhof and Todorov’s trustworthy and untrustworthy faces to investigate whether implicitly perceived trustworthiness shapes empathy.

Chapter 5

Experiment 4 and Experiment 5: Empathy and Trustworthiness

In Experiment 4 and Experiment 5 I investigated the impact of trustworthiness on neural empathic responses for pain. In Experiment 4 I adopted computerized Oosterhof and Todorov's trustworthy and untrustworthy faces; in Experiment 5 implemented an analogous design with a set of real faces that, in a first phase of the study, were rated on the dimension of trustworthiness by an independent sample.

As discussed at length in Chapter 1, a critical feature of empathy, as measured by both neuroimaging and event-related potentials (i.e., ERPs) techniques, is that it may be modulated by social and affective relations between individuals. As it might be expected, people tend to be more empathetic toward similar others − for instance in terms of group membership (see Bufalari & Ionta, 2013; Eres & Molenberghs, 2013) – an aspect that has been explored in Experiment 2.

In this vein, neural responses associated with empathy are shaped by learned preferences and appraisal of others' social behaviour. Singer and colleagues (2006) have offered an elegant demonstration of this kind of variance in empathy in the context of empathy towards others' pain, providing evidence that painfully stimulated unfair individuals induce in the observers reduced activations of empathy-related brain regions compared to fair individuals. Taken together, these findings very clearly confirmed the view that previous knowledge on someone else is a significant source of information that biases downstream processes, critically including processes empathy towards others' pain.

However, is previous knowledge on someone's fair/unfair social conduct an essential source of information for shaping people' empathic responses? Results as those reported in Experiment 2 strongly advocate that simply facial features diagnostic of a different group membership (i.e., skin colour) than that of the observer may modulate empathy. People indeed immediately form impressions of others on first meeting on the basis of others' physical appearance and immediately like or dislike them adjusting their behaviour even in the absence of previous knowledge on others' personality and social conduct. Evaluation of a stranger as trustworthy or untrustworthy is one of these appraisals 'at first sight' taking only a fraction of a second (e.g., Bar, Neta, & Linz, 2006; Todorov, Said, Oosterhof, & Engell, 2011; Willis & Todorov, 2006). Experiment 4 showed that trustworthiness appraisal is automatically performed (i.e., task-irrelevant) in less than half of a second following the presentation of a face that had to be memorized (see also Sessa, Tomelleri, Luria, Castelli, Reynolds, & Dell'Acqua, 2012). Importantly, this appraisal seems reliable: Convincing evidence substantiates that individuals perceived as untrustworthy tend in fact to exploit the trust of others in social and economic exchanges (Stirrat & Perrett, 2010); on the other hand, individuals perceived as trustworthy are more likely to possess a particular variation of the oxytocin receptor gene, known as the GG genotype, associated with a more prosocial and empathetic behavior (Kogan, Saslow, Impett, Oveis, Keltner, & Saturn, 2011).

The ongoing considerations led us to hypothesize that, even when information on others' social behaviour is lacking, empathy may be shaped solely by this first impression. In particular, we conjectured that individuals perceived as trustworthy would have induced in the observer a greater neural reaction to their pain when compared with individuals perceived as untrustworthy, similarly to individuals known to be fair/unfair.

By means of ERPs and source analyses techniques, I monitored neural reactions towards trustworthy and untrustworthy individuals in a painful context (i.e., a needle injection vs. a Q-tip touch) in two experiments using both computerized bald male faces parametrically manipulated on the trustworthiness dimension (Experiment 4) and real male faces rated on the trustworthiness dimension by an independent sample of participants (Experiment 5). Noteworthy, real faces do not have standardized facial features of trustworthiness as Oosterhof and Todorov's do so that more variability in trustworthiness appraisal might be expected. This second experiment on the effect of perceived trustworthiness in shaping an empathic reaction was relevant since a recent driving force in the social neuroscience field underlined the importance of the use of naturalistic stimuli in social cognition (Zaki & Ochsner 2009). Indeed, social neuroscientists address social issues without directly test cognitive processes in realistic social contexts; that might shape in an unpredictable way brain activity deployed to perform the experimental task. The authors highlighted that the processing of target states involve multimodal information that need to be dynamically integrated by perceivers. Further, external information is usually part of a social context that might constrain perceivers' interpretation of such information about targets' internal states.

As the reader might remember, previous work investigating the temporal aspects of empathy toward others' pain by means of the ERP approach consistently showed a positive shift of scalp-recorded electrical activity when participants were presented with painful stimulation applied to others relative to neutral stimulation in different temporal windows on the basis of the task at hand encompassing P2, N2, N3 and P3 ERP components. Experiment 1 built compelling and direct evidence of the relationship between early pre-P3 time-window, including P2, N2, N3, and mechanisms associated with experience sharing; and, by contrast, late time-window including the P3, which is instead associated with mechanisms underlying mentalizing. Critically, Fan and Han (2008) showed that the contextual reality of stimuli shapes electrophysiological empathic reaction by postponing it for cartooned stimuli relative to realistic stimuli.

In the present experiments, I implemented a variant of the pain decision task in which participants, in each experimental trial, were first exposed to a face looking either trustworthy or untrustworthy, and following a short blank interval, a syringe or a Q-tip were displayed, indicating that a painful stimulation or a non-painful stimulation, respectively, was applied to the presented face in that trial. The temporal separation between the face stimuli and the painful/non-painful stimuli was decisive in order to allow participants to implicitly appraise faces as trustworthy or untrustworthy before the presentation of the painful/non-painful stimuli. Participants were instructed to decide whether each face was painfully or neutrally stimulated by imaging that the stimulation associated with the object (i.e., syringe or Q-tip) was applied to the presented face.

Computerized faces with standardized facial features of trustworthiness were used in Experiment 4, whereas real and more ecologically valid faces, previously rated on the trustworthiness dimension, were used in Experiment 5.

5.1 Experiments 4: Empathy for computerized trustworthy and untrustworthy individuals' pain

5.1.1 Method

Participants.

Seventeen volunteer students (4 males; mean age: 23.16 years, SD = 2.48; 2 left-handed) participated in Experiment 4; all reported normal or corrected-to-normal vision and no history of neurological disorders; all gave their informed consent according to the ethical principles approved by the University of Padova.

Stimuli

It has been suggested that trustworthiness appraisal, in particular for those faces at the extremes of the trustworthiness dimension, approximates the detection of emotional facial expressions (e.g., Oosterhof & Todorov, 2008). In order to avoid potential confounds between emotion and trustworthiness appraisals in driving modulatory effects on empathy, similarly to Experiment 4, prior to the ERP experiment an independent student sample (N = 29; 6 males, mean age: 23.43 years, SD = 1.86) provided 7-step ratings of faces' trustworthiness (trustworthy vs. untrustworthy) and emotional expression (happy vs. angry) of 110 neutral bald Caucasian males generated using FaceGen Modeller 3.2 (Singular Inversions, 2007; +/-2 and +/-3 SD from neutral, in the middle of the trustworthiness dimension) according to the methods described by Oosterhof and Todorov (2008). The resulting 12 most trustworthy and 12 most untrustworthy faces with a non-significant correlation with the emotion scale (consensus neutral expression $p > .1$) were selected as face stimuli for the pain decision task.

Procedure

Pain decision task. Participants performed a modified version of the pain decision task. Each trial began with a randomly variable fixation cross at the center of the computer screen (800-1600 ms in steps of 100 ms). A trustworthy or untrustworthy face and either a Q-tip or a syringe evocating a neutral or a painful stimulation respectively, were interleaved by a variable blank interval (800-1600 ms in steps of 100 ms) and presented for 250 ms. The different types of trials were randomly intermixed. Participants were instructed to decide, without speed pressure, whether each face was painfully or neutrally stimulated by imaging that the stimulation evocated by the object was applied to the presented face. An example of face stimuli and a schematic illustration of the procedure used in Experiment 4 and Experiment 5 are depicted in Figure 1.

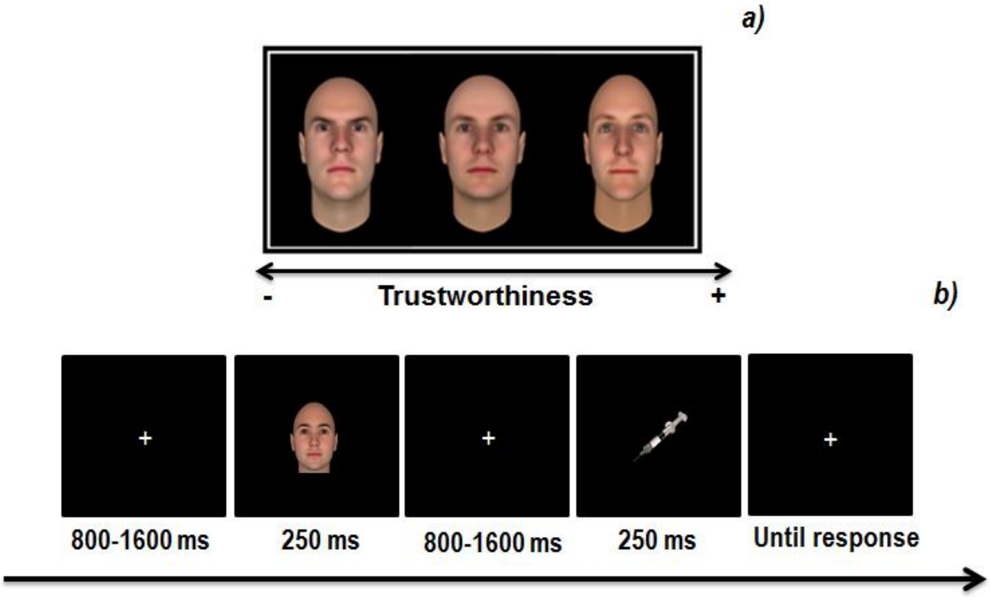


Figure 1. a) Example of computerized Oosterhof and Todorov’s trustworthy and untrustworthy faces, analogous as those used in Experiment 4. b) Schematic illustration of the modified version of the pain decision task. The face and the tool associated with either painful or neutral stimulation have been temporally separated in order to allow implicit trustworthiness appraisal.

Self-report questionnaires. Following the pain decision task, participants completed the Empathy Quotient (i.e., EQ) questionnaire developed by Baron-Cohen and Wheelwright (2004), and the Interpersonal Reactivity Index (i.e., IRI), which have been described in Chapter 1.

EEG acquisition and analysis. The EEG was recorded during the pain decision task from 64 active electrodes placed on an elastic Acti-Cap referenced to the left earlobe. The EEG was re-referenced offline to the average of the left and right earlobes. Horizontal EOG (HEOG) was recorded bipolarly from electrodes positioned lateral to the outer canthi of both eyes. Vertical EOG (VEOG) was recorded bipolarly from two electrodes, one above (Fp1) and one below the left eye. The impedance was kept less than 10KΩ. EEG, HEOG and VEOG signals was amplified (pass band 0.1−80 Hz) and digitized at a sampling rate of 250 Hz. The EEG was filtered before being analyzed (pass band 0.5–40 Hz and notch 50 Hz) and segmented into 1000-ms epochs starting 200 ms prior to the onset of the Q-tip/syringe. The epochs were baseline corrected based on the mean activity during the 200 ms prestimulus period, for each electrode site. Trials associated with incorrect responses or contaminated by large horizontal eye movements, eye blinks or other artifacts (exceeding ± 30μV, ± 60μV and ± 80μV respectively) were excluded from analysis. Following artifact rejection, separate average waveforms for each condition were generated time-locked to the Q-tip/syringe.

Mean ERPs amplitude values were measured at pooled electrodes selected from the fronto-central (Fz, FCz, F1–F2, F3–F4, FC1–FC2, FC3–FC4), and centro-parietal (CPz, CP1–CP2, CP3–CP4, Pz, P1-P2, P3-P4) regions, in time windows of 400-500 ms and of 300-500 ms. In all multi-factorial analyses, a Greenhouse-Geisser correction was used where appropriate.

Neural reaction to pain was calculated by subtracting mean amplitude values for neutral stimulation from those registered for painful stimulation.

The standardized Low Resolution Brain Electromagnetic Tomography (sLORETA) (Pascual-Marqui, 2002) was used for brain localization of the potential sources of ERP reactions to pain. sLORETA extrapolates 3D statistical maps from EEG data of the possible sources of scalp-recorded ERP components. Using sLORETA, 3D maps and stereotaxic information about current density source of neural activity modulated by the painful vs. non-painful manipulation were derived from the present EEG dataset. The analysis was conducted following the creation of a boundary element method (BEM) model, including cortical and skin, with about 5000 nodes from magnetic resonance imaging (MRI) data, the selection of a temporal window in which P3 ERP responses differentiated between painful and non-painful stimulations (i.e., P3 reactions to pain), and a location-wise inverse weighting from the Minimum Norm Least Square (MNLS) analysis with estimated variances.

5.1.2 Results

Pain decision task.

Individual mean proportions of correct responses and reaction times (RTs) were submitted to a repeated measure analysis of variance (ANOVA) considering stimulation condition (painful vs. non-painful) and trustworthiness (trustworthy faces vs. untrustworthy faces) as within-subjects factors. ANOVA carried out on mean proportions of correct responses did not show any significant effects ($Fs < 1$). ANOVA carried out on reaction times showed a significant effect of the interaction $F(1,16) = 5.291$, $p = .035$; $\eta_p^2 = .249$ indicating that participants were more responsive to the painful stimulation (414 ms) relative to the neutral stimulation (452 ms) applied to trustworthy faces ($t(16) = -2.059$, $p = .056$).

EQ. Mean EQ score was similar to those found in the original study (Baron-Cohen and Wheelwright, 2004), i.e. 42.12 (*SD* = 6.43).

IRI. Scores were computed by summing 1–7 scores to each item of the four subscales. Inter-individual mean rating scores were 3.61 (*SD* = 0.48) for the PT subscale, 3.4 (*SD* = 0.57) for the FS subscale, 3.71 (*SD* = 0.36) for the EC subscale, and 2.91 (*SD* = 0.58) for the PD subscale.

ERPs: P3

Visual inspection of electrophysiological results shows effects on the P3 compoenent. This is not surprising as the modified version of the pain decision task used in Experiment 4 needed imagination to be performed. Goldman and Jordan (2013) refer to imagination as a high-level process of mentalizing. Indeed, it specifically required participants to form an inner representation of the other's pain.

Figure 2a shows P3 waveforms time-locked to syringe/Q-tip elicited at fronto-central and centro-parietal sites in painful and non-painful conditions and separately for trustworthy and untrustworthy faces. Statistical analyses on the fronto-central and centro-parietal P3 mean amplitude values were submitted to a 2 (stimulation: painful vs. non-painful) x 2 (trustworthiness: trustworthy vs. untrustworthy faces) repeated measures ANOVA as within-subjects factors. The centro-parietal P3 was larger for painful stimuli than non-painful stimuli, $F(1, 16) = 20.685$, $p < .001$; $\eta_p^2 > .564$). Most importantly, ANOVAs showed significant interactions between stimulation and trustworthiness on both the fronto-central P3 ($F(1, 16) = 16.397$, $p < .005$; $\eta_p^2 > .506$) and on the centro-parietal P3, $F(1,16) = 8.007$, $p < .05$; $\eta_p^2 > .334$. Planned comparisons indicated that painful stimuli elicited a positive shift of the fronto-central P3 relative to non-painful stimuli only when applied to trustworthy faces ($t(16)$ =, p < .05); this shift of the fronto-central P3 was absent when untrustworthy faces were painfully stimulated (*t* < 1).

The results of sLORETA analysis revealed that the magnified P3 reaction to the pain of trustworthy individuals compared to the pain of untrustworthy individuals − i.e., [trustworthy faces (painful minus non-painful stimulation conditions)] minus [untrustworthy faces (painful minus non-painful stimulation conditions) − was mainly localized in the precuneus (peak Montreal Neurological Institute, i.e. MNI, coordinates: 5, -55, 65) along with the superior temporal gyrus (peak MNI coordinates: 45, 25, -20), the middle temporal gyrus (peak MNI coordinates: 55, 10, -20), a portion of the inferior frontal gyrus (Brodmann Area, BA, 47; peak MNI coordinates: 40, 30, -20), the supramarginal gyrus (peak MNI coordinates: -55, -45, 35) and the inferior parietal lobule (peak MNI coordinates: -55, -35, 35). Since almost all of these brain regions are involved in the attribution of mental states to others and intentional empathy (e.g., Amodio and Frith, 2006; de Greck, Wang, Yang, Wang, Northoff, and Han, 2012; Frith and Frith, 1999; Moriguchi, Ohnishi et al., 2006; Saxe & Kanwisher, 2003; Spiers & Maguire, 2006), these results suggest that participants devolved a larger amount of cognitive resources while mentalizing on the pain of trustworthy individuals compared to untrustworthy individuals (see Figure 2b).

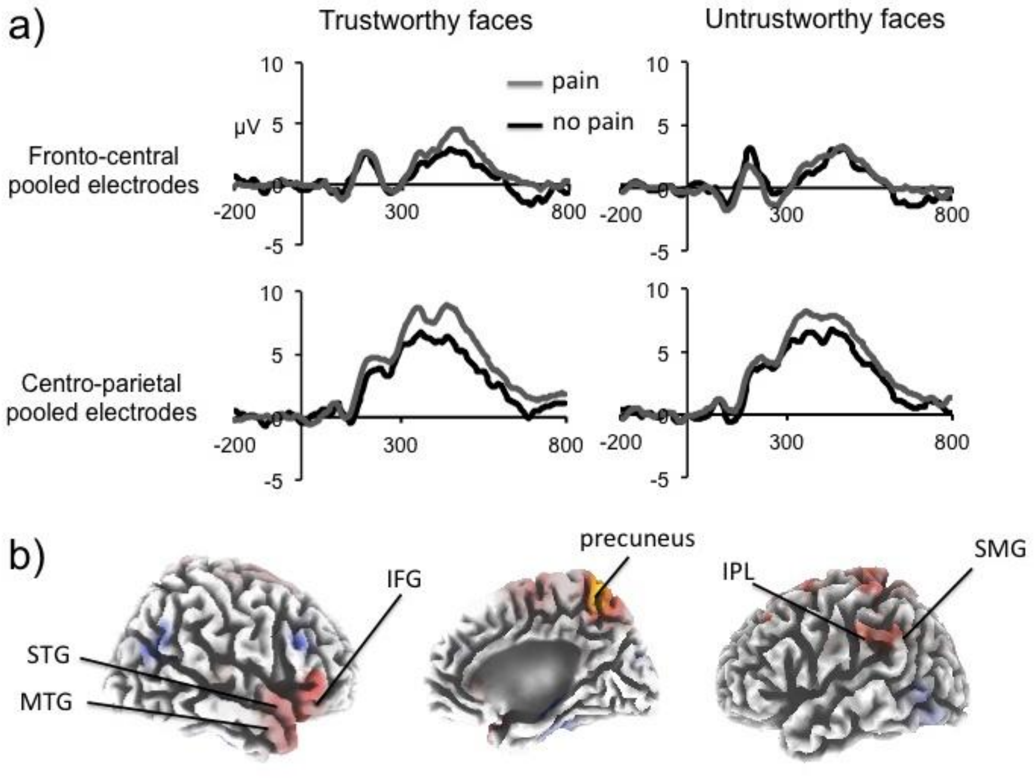


Figure 2. a) ERPs reactions to painful (i.e. grey lines) and neutral (i.e. black lines) for trustworthy (i.e. right panels) and untrustworthy (i.e. left panels) registered at pooled fronto-central electrodes (i.e. upper panels) and at pooled centro-parietal electrodes (i.e. lower panels). b) Sources estimation of the selective response to the pain of trustworthy-looking individuals (compared to untrustworthy-looking individuals; i.e., trustworthy faces [painful minus non-painful stimulation conditions] minus untrustworthy faces [painful minus non-painful stimulation conditions]). Precuneus (peak MNI coordinates: 5, -55, 65); STG = superior temporal gyrus (peak MNI coordinates: 45, 25, -20); MTG = middle temporal gyrus (peak MNI coordinates: 55, 10, -20); IFG = inferior frontal gyrus (peak MNI coordinates: 40, 30, -20); SMG = supramarginal gyrus (peak MNI coordinates: -55, -45, 35); IPL = inferior parietal lobule (peak MNI coordinates: -55, -35, 35).

I also computed pain reactions in the P3 time-window for both the fronto-central and the centro-parietal P3 by subtracting P3 mean amplitude values elicited in the non-painful stimulation condition from P3 mean amplitude values elicited in the painful condition, separately for trustworthy faces and untrustworthy faces. No correlations have been registered with individual IRI subscale scores. By contrasts, the centro-parietal P3 reactions to pain correlated with the EQ scores when trustworthy faces were presented, *r* = 5.28, *p* < .05. This correlation was also marginally significant for untrustworthy faces, *r* = 3.43, *p* = .089. These findings support the conclusion that the more pronounced P3 reaction to the pain of trustworthy individuals indexes indeed a magnified empathic response. Figure 3 shows the scatterplot of the correlation between individual EQ scores and P3 pain reactions (i.e., painful minus non-painful conditions) recorded at centro-parietal electrode sites towards trustworthy faces.

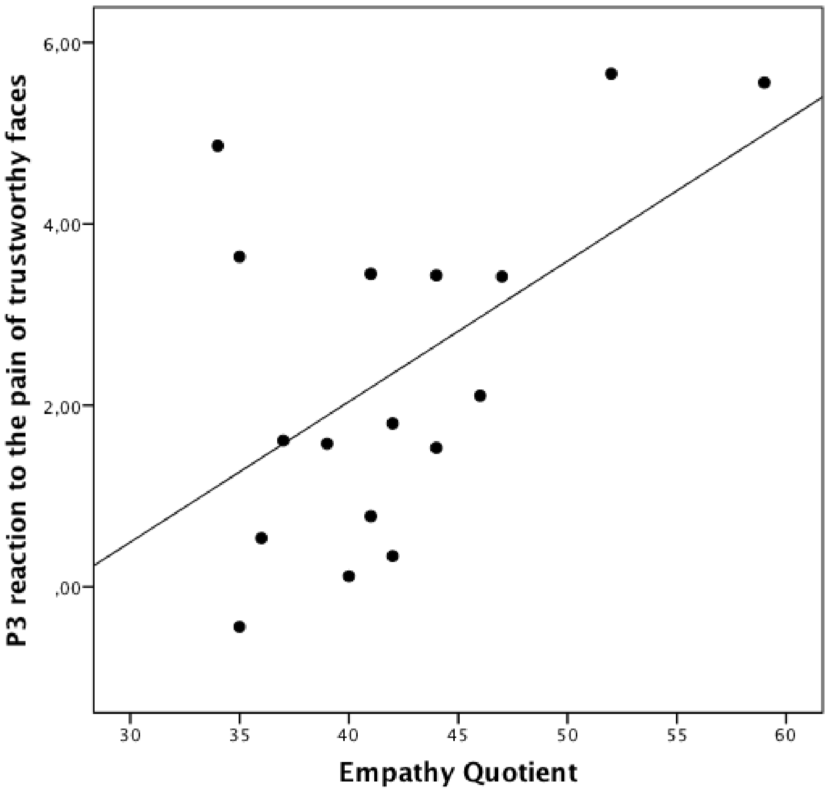


Figure 3. Scatterplot of the correlation between individual EQ scores and P3 pain reactions (i.e., painful minus non-painful conditions) towards trustworthy faces.

5.1.3 Discussion

P3 ERP reactions to the pain of individuals looking trustworthy recorded at fronto-central and centro-parietal electrode sites were magnified relative to P3 reactions to the pain of individuals looking untrustworthy. Following previous studies using similar experimental paradigms (Avenanti et al., 2010; Decety et al., 2010; Fan & Han, 2008; Li & Han, 2010; Xu et al., 2009), I interpret these P3 reactions to pain as reflecting the engagement of empathy-related processes. The positive correlation between such P3 ERP reactions recorded at centro-parietal electrode sites and the EQ scores can be taken as evidence compatible with this view. More specifically, and in line with results of Experiment 1 and of others scholars (e.g., Decety et al., 2010; Fan & Han, 2008), I suggest that the P3 reactions to pain primarily index mentalizing. Convergent with this interpretation, the likely source of the more pronounced P3 reaction to the pain of trustworthy (vs. untrustworthy) individuals was localized in the precuneus, a key region of the mentalizing network (e.g., Lamm et al., 2011; Van Overwalle & Baetens, 2009; Zaki & Ochsner, 2012) with a well-established critical role in social cognition tasks (e.g., Harris, Todorov & Fiske, 2005).

5.2. Experiment 5: Empathy for trustworthy and untrustworthy individuals' pain − real faces

5.2.1 Method

Participants

Seventeen volunteers students (4 males; mean age: 22.35 years, SD = 3.04; 4 left handed) participated in Experiment 5; all reported normal or corrected-to-normal vision and no history of neurological disorders; all gave their informed consent according to the ethical principles approved by the University of Padova.

Stimuli

It has been underlined the importance of the naturalism in social neuroscience studies (Zaki & Ochsner, 2009). In order to investigate modulatory effects on empathy in a more ecological context, prior to the ERP experiment an independent students sample (N = 45; 6 males, mean age: 23.48 years, SD = 2.3) provided 7-step ratings of faces' trustworthiness (trustworthy vs. untrustworthy) of 50 real Caucasian faces with standardized neutral expression included in Eberhardt's database. The resulting 10 most trustworthy and 10 most untrustworthy faces were selected as stimuli for the pain decision task.

Figure 4. Example of real trustworthy and untrustworthy faces.

Procedure

The procedure was the same as in Experiment 4. The only difference was that the backcolor of the screen was light gray instead of black.

EQ. Following the pain decision task, participants completed the Empathy Quotient (i.e., EQ) questionnaire developed by Baron-Cohen and Wheelwright (2004), which has been described in Chapter 1.

EEG acquisition and analysis. The EEG was recorded during the pain decision task with same parameters as those used for Experiment 4. The EEG was filtered before being analyzed (pass band 0.5–40 Hz and notch 50 Hz) and segmented into 1000-ms epochs starting 100 ms prior to the onset of the Q-tip/syringe The epochs were baseline corrected based on the mean activity during the 100 ms prestimulus period, for each electrode site. Trials associated with incorrect responses or contaminated by large horizontal eye movements, eye blinks or other artifacts (exceeding ± 30μV, ± 60μV and ± 80μV respectively) were excluded from analysis. Following artifact rejection, separate average waveforms for each condition were generated time-locked to the Q-tip/syringe.

Mean ERPs amplitude values were measured at the same fronto-central and centro-parietal electrodes as in Experiment 4, in time-windows of 320–450 ms and 300–980 ms. In all multi-factorial analyses, a Greenhouse-Geisser correction was used where appropriate.

Neural reaction to pain was calculated by subtracting mean amplitude values for neutral stimulation from those registered for painful stimulation

5.2.2 Results

Pain decision task. Individual mean proportions of correct responses and reaction times were submitted to a repeated measure analysis of variance (ANOVA) considering stimulation condition (painful vs. non-painful) and trustworthiness (trustworthy faces vs. untrustworthy faces) as within-subjects factors. ANOVAs on both accuracy and reaction times did not show significant main effects nor did interaction (all *Fs* < 1).

EQ. Mean EQ score was similar to those found in the original study (Baron-Cohen & Wheelwright, 2004), i.e. 48.65 (*SD* = 10.75).

ERPs.

Visual inspection of electrophysiological results shows effects in an earlier time-window (i.e., N2-N3 time-range) relative to those observed in Experiment 4, although it is sustained until the P3 component time-range. This is might be due to the naturalism of the stimuli used.

Figure 4 shows N2-N3 and P3 waveforms time-locked to syringe/Q-tip elicited at fronto-central and centro-parietal sites in painful and non-painful conditions and separately for trustworthy and untrustworthy faces. Statistical analysis on the front-central and centro-parietal ERPs mean amplitude values were submitted to a 2 (stimulation: painful vs. non-painful) x 2 (trustworthiness: trustworthy vs. untrustworthy faces) repeated measures ANOVA as within-subject factors for both the N2-N3 activity (320-450 ms) and the P3 (300-980 ms) for fronto-central and centro-parietal pooled electrodes, separately.

N2-N3. The fronto-central N2-N3 activity showed a positive shift for painful compared to non-painful stimuli selective for trustworthy faces: the interaction between stimulation and trustworthiness was significant, $F(1,16) = 11.285$, $p <.005$; $\eta_p^2 = .414$. Post-hoc analysis showed significant comparison between painful and non-painful stimuli for trustworthy faces ($t\ (16) = 3.071$ $p = .007$) but not for untrustworthy faces ($t < 1$).

The centro-parietal N2-N3 activity was larger for painful (4.793 μV) relative to neutral (3.580 μV) stimulation as indicated by the main effect of stimulation $F(1,16) = 17.743$, $p = .001$; $\eta_p^2 = .526$. Most importantly, the ANOVA showed significant interaction between stimulation and trustworthiness, $F(1,16) = 6.338$, $p = .023$; $\eta_p^2 = 0.284$. Post-hoc analysis indicated that painful stimulation elicited a positive shift of the centro-parietal N2-N3 relative to non-painful stimuli only participants were required to imagine that they were applied to trustworthy faces ($t(16) = 4.215$, $p = .001$); this shift of the centro-parietal N2-N3 was absent for untrustworthy ($t\ (16) = 1.823$ $p = .087$).

P3. The fronto-central P3 was more positive for painful (-0.722 μV) relative to neutral (-1.090 μV) stimulation as indicated by the main effect of stimulation $F(1,16) = 4.570$ $p < .05$; $\eta_p^2 = .222$.

The centro-parietal P3 was more positive for painful (2.245 μV) relative to neutral (1.448 μV) stimulation as indicated by the main effect of stimulation $F(1,16) = 18.116$ $p = .001$; $\eta_p^2 = .531$. Most importantly, the ANOVA showed a significant interaction between stimulation and trustworthiness, $F(1,16) = 5.335$, $p = .035$, $\eta_p^2 = 0.250$. Post-hoc analyses indicated that painful stimulation elicited a positive shift of both the fronto-central and the centro-parietal P3 relative to non-painful stimuli only when applied to trustworthy faces (min $t(16) = 2.328$, max $p = .033$); this shift was absent for untrustworthy faces (max $t(16) = 1.566$, min $p = .135$).

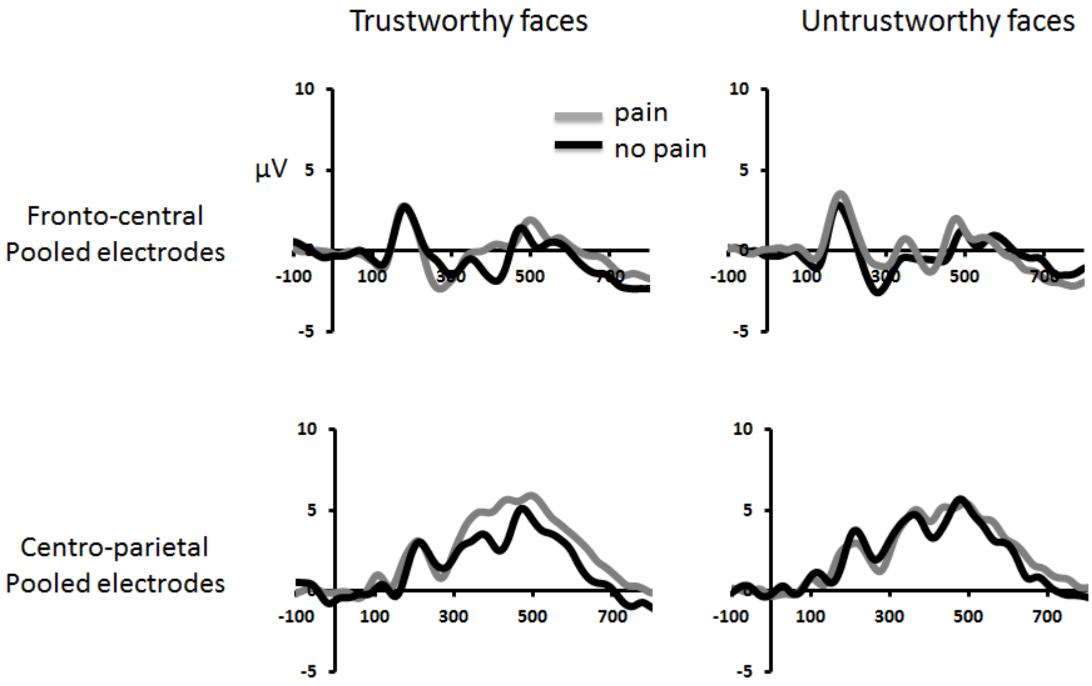


Figure 4. ERPs reactions to painful (i.e. grey lines) and neutral (i.e. black lines) for trustworthy (i.e. right panels) and untrustworthy (i.e. left panels) registered at pooled fronto-central electrodes (i.e. upper panels) and at pooled centro-parietal electrodes (i.e. lower panels).

Crucially, the reaction to the pain of real trustworthy individuals was shifted in an earlier temporal window compared to that observed in Experiment 4.

I also computed pain reactions in the N2-N3 and the P3 time-windows for both the fronto-central and the centro-parietal ERPs by subtracting ERPs mean amplitude values elicited in the non-painful stimulation condition from ERPs mean amplitude values elicited in the painful condition, separately for trustworthy faces and untrustworthy faces. No correlations have been registered with individual EQ scores. By contrasts, either the fronto-central and the centro-parietal N2-N3 reaction to pain positively correlated with both the Perspective Taking (i.e., PT; $r = .496$, $p = .021$ at fronto-central sites; $r = .586$, $p = .007$ at centro-parietal sites) and the Empathic Concern (i.e., EC; $r = .539$, $p = .013$ at fronto-central sites; $r = .551$, $p = .011$ at centro-parietal sites) subscale scores of the IRI when untrustworthy faces were presented, see Figure 5.

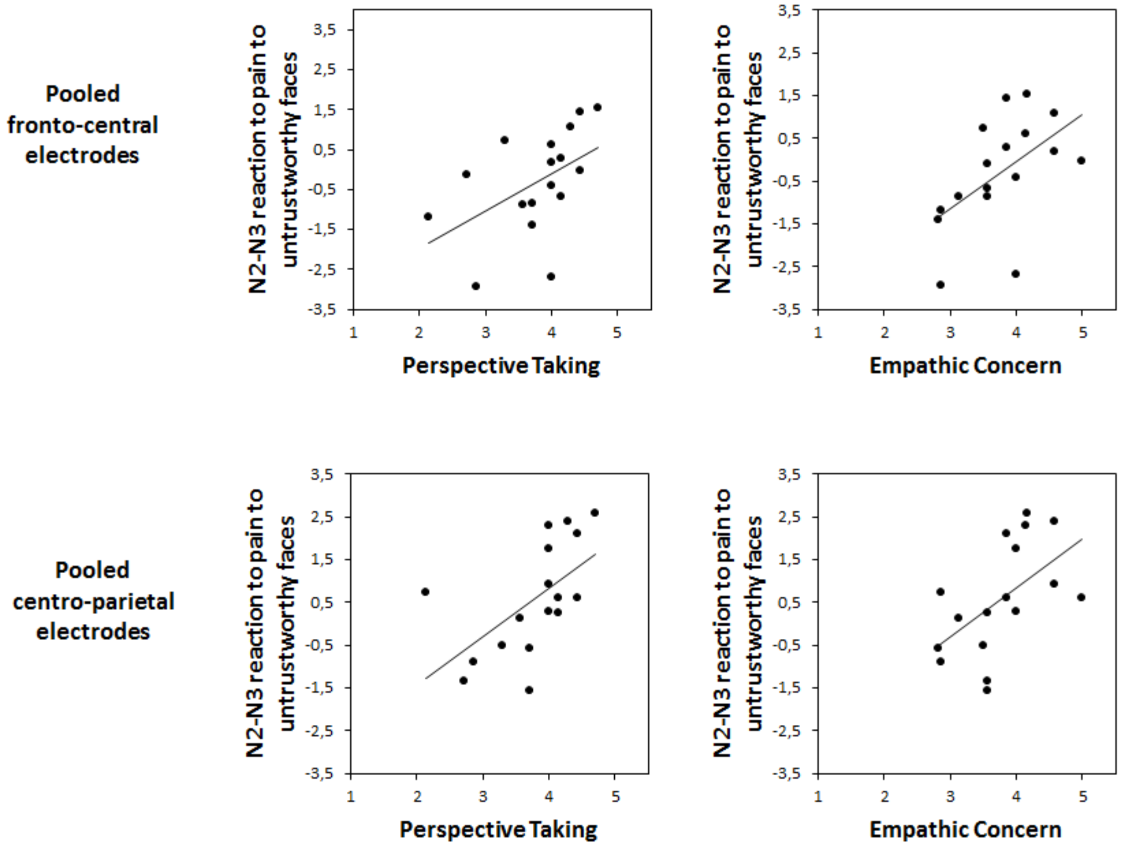


Figure 5. Scatterplot of the correlation between individual Perspective Taking and Empathic Concern subscale scores and N2-N3 pain reactions (i.e., painful minus non-painful conditions) towards untrustworthy faces recorded at fronto-central and at centro-parietal pooled electrode sites, separately.

Although such correlation was observed only for untrustworthy faces' pain but was not significant for trustworthy faces' pain, it further strengthens the association between the N2-N3 differential activity and an empathic reaction to others' pain. Interestingly, the Pt and the Ec subscales are related to the two subprocesses of empathy, i.e. mentalizing and experience sharing, respectively. This is in not surprising as the specific version of the pain decision task and the ecological validity of the stimuli have been used in the present experiment. Indeed, the experimental paradigm specifically required participants to imagine that either a painful or neutral stimulation was applied to the presented face. At one hand, similarly to Experiment 4, it engages necessarily mentalizing mechanisms. On the other hand, in line with Zaki and Ochsner's observations (2009) about the necessity of ecological validity in social cognition, real faces might involve, relative to computerized faces, more automatic and affective processes, and so partly engaging also experience sharing mechanisms.

5.2.3 Discussion

N2-N3 and P3 ERP reactions to pain were magnified relative to P3 reactions to the pain of individuals looking untrustworthy. Following Experiment 4, I interpret these ERP reactions to pain as reflecting the engagement of empathy-related processes. Noteworthy, relative to results of Experiment 4, this reaction to pain was observed in an earlier temporal window that, in line with the results of Experiment 1 and other scholars (e.g., Decety et al., 2010; Fan & Han, 2008), reflects the engagement of experience sharing. Importantly, neural reaction to pain positively correlated with individual empathy traits, supporting its functional meaning as related to empathy.

5.3 General Discussion

Both computerized and real trustworthy faces elicited magnified neural responses to pain when compared to untrustworthy faces as observed on both the P3 (Experiment 4) and the N2-N3 and the P3 (Experiment 5). I interpreted these ERP reactions to pain as reflecting the engagement of empathy-related processes. The correlations between P3 and N2-N3 ERP reactions recorded and the EQ scores can be taken as evidence compatible with this view.

At least in Experiment 4, the likely neural source of the P3 empathic reaction to pain was the precuneus, a key region of the mentalizing system. These findings dovetail nicely with previous evidence showing that the precuneus is activated when adopting the perspective of another person to imagine/understand his or her emotional reactions or pain (Ruby & Decety, 2004; Jackson, Meltzoff, & Decety, 2006) or when perceiving faces of significant others (Gobbini, Leibenluft, Santiago, & Haxby, 2004).

Furthermore, greater activity in the precuneus (along with medial prefrontal cortex, bilateral temporo-parietal junction and posterior cingulate cortex) has been observed when members of the ingroup were presented in conditions of emotional pain compared to when members of the outgroup were presented in similar painful conditions (Cheon, Im, Harada, Kim, Mathur, Scimeca, Parrish, Park, & Chiao, 2011) leading to the fascinating hypothesis that trustworthy individuals are perceived as more proximate to oneself. Complementary evidence to this proposal is offered by an fMRI study that investigated differences in individuals' judgments on physical similarity to the self of morphs created by mixing the self-face with novel trustworthy or untrustworthy faces (Verosky & Todorov, 2010). Notably, Verosky and Todorov reported that participants were more likely to identify the morphs with trustworthy faces as looking like the self, compared to morphs with untrustworthy faces. In addition, activity in several brain regions differentiated between the self and untrustworthy faces to a much greater extent than between the self and trustworthy faces. Although the following may be considered just a speculation, these findings might support the view that magnified P3 reactions observed in the present study to the pain of individuals looking trustworthy (compared to the pain of individuals looking untrustworthy) originated from the fast implicit appraisal of these individuals as more closely linked to the self. If this interpretation is correct, greater empathic responses towards individuals looking trustworthy (vs. individuals looking untrustworthy) are analogous to the greater empathic responses previously observed towards members of the ingroup (vs. members of the outgroup) and individuals in close affective relationships (vs. strangers).

Remarkably, results of Experiment 4 with computerized faces have been replicated with ecologically valid real faces consensually looking as either trustworthy or untrustworthy (Experiment 5). Indeed, neural reactions to pain observed in real faces have been observed in an earlier time-window, pre-P3, that is related to mechanisms underlying experience sharing, as demonstrated, for instance, in Experiment 1.

That is, the naturalism of the stimuli affected empathy in a different fashion. Computerized faces induced magnified P3 reaction to pain, involving mentalizing mechanisms, whereas real faces induced magnified reaction to pain also in an earlier time-window pre-P3, namely in the N2-N3 time-window, involving experience sharing mechanisms. Consistently with these findings, Fan and Han (2008) presented participants with either realistic or cartooned hands (i.e., one or two hands) that could be either painfully or neutrally stimulated. Participants underwent in the same experimental session both a pain decision task and a counting task[16]. Results showed that cartoons elicited a later ERP reaction to pain relative to real stimuli, indicating that the contextual reality of stimuli postponed the early empathic processing of pain.

Consistently with this view, Zaki and Ochsner (2009) highlighted that realistic social contexts might shape in an unpredictable way brain activity deployed to perform the experimental task because it involves multimodal information that need to be dynamically integrated by perceivers. The results of the Experiment 5 are in line with those of Spunt and Lieberman (2011), who suggested that the connection between mirror and mentalizing that they found in support of an integrative model of empathy, might have been conveyed by the naturalism of their stimuli.

In conclusion, I observed decreased empathy-related neural responses toward untrustworthy-looking individuals compared to trustworthy-looking individuals − likely related to reduced activation of brain regions previously associated with mentalizing and intentional empathy − providing, to my knowledge, the first proof that empathy may be shaped by first impressions of others. Although this sensitivity of empathy to others' physical features may be adaptive in a variety of social contexts, especially

---

[16] The counting task required participants to say the number of hands presented: either one or two.

when considering its established reliability (e.g., Kogan et al., 2011; Stirrat & Perrett, 2010), the present findings may be of particular importance in other delicate contexts where looking untrustworthy may be detrimental, such as legal processes or care-giving. Importantly, in this last context, for instance, it has been recognized that showing empathy to patients, whatever their physical appearance, can improve care (Canadian Medical Association Journal, 2011).

Chapter 6

Conclusions

The present thesis described a series of five ERP experiments aimed at exploring the multifaceted nature of empathy (Chapter 2; Experiment 1) and the variance in empathic processes, as a function of others' race (Chapter 3; Experiment 2) and others' perceived trustworthiness (Chapter 4 and Chapter 5; Experiment 3, 4 and 5) of a face, addressed through classical and modified versions of the pain decision task.

Dissecting multifaceted empathy. In Experiment 1, I directly tested the hypothesis according to which the existing evidence of an anatomical dissociation between experience sharing and mentalizing is followed by complementing functional dissociation, or rather interaction, in the temporal domain. That was the first aim of the present studies. To address this issue, I implemented a novel version of the pain decision task in which both sensorimotor (picture of a face with either painful or neutral expression) and contextual cue (a sentence describing either a painful or neutral context) were presented.

One of the pitfalls of research on empathy concerns the oversimplification of stimuli. Of course, it has been necessary to isolate as much as possible elements that could enhance only one of the subprocesses of empathy in order to build rigorous knowledge about the basic structure of empathy. But consequently, at present, compelling scientific, yet epistemological, evidence either supporting or rejecting an integrative model of empathy is still under examination. In an attempt to make a step forward in this direction, the concurrent presence of both types of cues was necessary in order to uncover the specific contribution of each cue and to allow the observation of an interaction between them, if present. Fan and Han (2008) provided the first suggestion that there is a family of ERP components that are modulated by bottom-up mechanisms in an early time window pre-P3, including P2, N2 and N3; and a later time-window, including the P3 component that is modulated by top-down mechanisms.

In line with this first evidence, the results of Experiment 1 showed that experience sharing and mentalizing are, to a large extent, dissociable mechanisms.

In the time-window including the P2 and the N2-N3 the *perceptual cue* selectively activated mechanisms underlying the experience sharing; in the immediately following time-window, which includes the P3, the *contextual cue* selectively activated mechanisms underlying the mentalizing. In support of such view, empathic reaction to pain registered in these time-windows have been associated with empathic abilities underlying experience sharing, the earlier, and mentalizing, the latter.

However, an intermediate time-window between N2-N3 activity and the P3 peak, was sensitive to both sensorimotor and contextual cues. Crucially, the concurrent presence of pain of both sensorimotor and contextual cues did not magnify empathic reaction to pain, it is rather the result of the contribute of the pain of only one of the cues. This might be an index of an unspecific response, probably not strictly empathy-related that mediates the transfer of information between the two subprocesses. Of course, future studies should investigate more directly such hypothesis.

Through the temporal deployment of electrophysiological reaction to pain, I highlighted how the subprocesses functionally contribute at any given time by dissecting empathy into two time-windows, one earlier, pre-P3 associated with mechanisms underlying experience sharing, and one later, that includes the P3 component, associated with cognitive mentalizing.

In the light of this new evidence, I explored the second aim of the present studies: Variance in neural empathic response to pain as function of race and perceived trustworthiness of a face.

Variance in empathic processes: Race. Complementing existing literature (e.g., Avenanti et al., 2010; Xu et al., 2009), I demonstrated with Experiment 2 that whereas experience sharing is modulated by race, mentalizing is not affected by it. Neuroscientific studies suggested that people are more naturally empathic towards own-race individuals relative to other-race individuals (e.g., Xu et al., 2009) and that this empathic bias is dependent on individual differences in implicit racial prejudice (Avenanti et al., 2010; Azevedo et al., 2012; Forgiarini et al., 2011). Noteworthy, the neuroscientific technique involved in these previous studies, e.g. fMRI and TMS, are suboptimal to clearly define whether this bias is related to all empathic processes or rather to an empathy subprocess. By virtue of the excellent temporal resolution of the ERPs, I explored the time-course of the empathic response to pain in White participants, which were exposed to own- and other-race faces either penetrated by a needle or touched by a Q-tip.

Results revealed that implicit racial bias is confined to experience sharing but empathic response towards other-race pain is observable in a later time-window linked to mentalizing.

Specifically, I observed a positive shift in the N2-N3 time range when White participants were exposed to White individuals in a painful condition but not when Black individuals were in in a painful condition. sLORETA identified the IFG (BA 45) as the neural source of this empathic reaction. This region of the IFG is known to be a core region of the putative mirror neuron system; a large body of evidence indicates that it is involved in the recognition of others' emotions (e.g., Chakrabarti, Bullmore, & Baron-Cohen, 2006; Frühholz & Grandjean, 2012) and similarly to the ACC, it concurs to the experience sharing component of empathy (e.g., Chakrabarti et al., 2006; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009). In addition, this difference was associated with individual EC scores of the IRI.

Furthermore, I observed an empathic reaction in the time-window including the P3 component. Notably, it was of comparable size between own- and other-race faces. Nevertheless, source localization of this later ERP response provided different source estimates indicating that empathic response to own- and other-race pain were qualitatively different. That is, the left TPJ (BA 40; involved in in self/other distinctions and autobiographical memories, Frith & Frith, 2001) was estimated as the neural source of the P3 empathic reaction to other-race pain; the left MFG (BA 9; involved in mentalizing; Spiers & Maguire, 2006) was estimated as the neural source for the reaction to own-race pain.

In this study only a weak association has been found with implicit racial prejudice (i.e., the empathic reaction to own-race faces was magnified for higher pro-White attitude, as measured by the IAT). Indeed, whereas implicit racial prejudice has been shown to be directly associated with the corticospinal inhibition (Avenanti et al., 2010), the SCR (Forgiarini et al., 2012) and brain areas involved in experience sharing (Azevedo et al., 2012), it did not affect scalp electrical activity (i.e. see also Sheng & Han, 2012) or at least not pervasively (Experiment 2). Mathur and collegues (Mathur, Harada, Lipke & Chiao, 2010) showed that the implicit racial prejudice was not associated with brain areas involved in cognitive empathy, other studies did not even measured it (e.g. Cheon et al., 2012; Xu et al., 2009). I propose that the interplay between cross-racial empathy for pain and implicit racial prejudice has not been exhaustively disclosed. Thus, social psychology reports that the implicit racial bias can be expressed in terms of prejudice and stereotype, which are related to either affective or semantic/cognitive mechanisms, respectively (Amodio & Ratner, 2011). Furthermore, prejudice and stereotype can be measured separately and independently by two different types of IAT, which have been specifically implemented and validated to measure either the prejudice and the stereotype (Amodio & Devine, 2006). In all the described studies investigating cross-racial empathy for pain, only the prejudice race-IAT has been measured. This observation opens new theoretically relevant scenarios. For instance, semantic integration starts early in processing, within 200 ms post-stimulus onset (e.g. Penolazzi, Hauk, & Pulvermüller, 2007) and can be observed in a quite wide time-ranges, from the N400 (e.g., semantic violation-related ERP component; e.g. Kutas & Hillyard, 1983) to the P600 (e.g. van Herten, Kolk, & Chwilla, 2005). It has recently been shown that semantic relations between words can modulate an earlier ERP component, namely the N2pc, which is a lateralized component, recorded at posterior electrode sites, which reflects attentional suppression of distractors stimuli during target(s) processing (e.g. Luck & Hillyard, 1994; Dell'Acqua, Sessa, Toffanin, Luria & Jolicoeur, 2010), or the enhancement of target(s) processing among distractors (Eimer, 1996). A semantic relationship between target and distractors reduces the amplitude of the N2pc starting at 170-180 ms post-stimulus onset (Dell'Acqua, Pesciarelli, Jolicoeur, Eimer & Peressotti, 2007). This finding provided evidence that cognitive effects such those related to semantic associations, can trigger very early modulations measurable online by using ERPs. On the basis of these considerations, it is worthwhile wondering whether either earlier or later scalp electrical activity as well as fMRI signal can be modulated directly by semantic associations with racial stereotype (assessed through the stereotype-IAT; Amodio & Devine, 2006) clarifying whether or not individual differences in implicit racial bias can be associated, to some extent, with empathic reaction to other-race pain.

Indeed, when an observer is exposed to a face of a different race, this immediately influences observer's physiological responses, likely mediated by the activity of the amygdala, which conveys very fast responses. For instance, when White participants are exposed to pictures of Black people, they show differences in muscular activity, skin conductance, startle-eyeblink, cardiovascular response (see Eberhardt, 2005 for a review), and ERPs associated not only to VWM resolution but also to the perception (Ofan, Rubin & Amodio, 2011) and attention (Amodio, 2010; Dickter & Bartholow, 2007; Ito & Urland, 2003). Interestingly, Sessa et al., (2012) conducted an ERP experiment that provided evidence of a link between the resolution of visual representations of other-race faces in VWM − as reflected by the Sustained Posterior Contralateral Negativity (i.e., SPCN) − and participants' implicit racial prejudice assessed through the prejudice-race IAT (i.e., the standard-race IAT), such that higher prejudiced participants encoded in VWM lower-resolution representations of other-race faces. This evidence suggests that, although more sensitive to individual differences of implicit racial prejudice, VWM representations of own- and other-race faces are maintained with different resolutions so that race influences VWM.. Experiment 2 showed that similarly, race influences also higher order cognitive processes, such as those implied in social interaction, namely empathy to others' pain.

Variance in empathic processes: Trustworthiness. Although implicitly appraised, race of a face is processed quickly and automatically, driven by physical facial features. Recently it has been demonstrated that facial trustworthiness is appraised at first sight in a fraction of second (Willis and Todorov, 2006), similarly to race, at least when explicitly requested. In Experiment 3 I tested the trustworthiness evaluation by using standardized trustworthy and untrustworthy faces such as Oosterhof and

Todorov's computerized 2D models, to address whether physical facial features of trustworthiness are automatically represented in VWM even when irrelevant for the task. Such a finding would strengthen the relevance of trustworthiness dimension in the first impression people form when encountering strangers.

To this aim, I adopted the change detection task, which has been specifically designed to test VWM representations. In this task, participants are usually required to memorize the identity of the face presented in the visual hemifield previously cued by an arrow and to ignore the one presented in the non-cued hemifield; after a brief blank interval, the faces were presented again and participants were required to say whether the test and the memorized faces were identical or not (see Sessa et al., 2010 for similar paradigm). I monitored the SPCN, which is an electrophysiological marker of the quantity (e.g. Vogel and Machizawa, 2004), the quality (e.g. Luria et al., 2010) and the resolution (e.g. Sessa et al., 2011) of visual information maintained in memory at any given moment during the retention interval, when the visual information is held in memory and before a response is required. SPCN mean amplitudes showed that trustworthiness is extracted from faces, even when task-irrelevant, modulating VWM representations of faces such as trustworthy faces are represented with greater resolution relative to untrustworthy faces. Interestingly, this evidence depends on individual differences of participants' general and social anxiety. This result suggests that although conveyed in a bottom-up manner, physical facial features of trustworthiness can also be subjective to individual differences in valence attribution/interpretation. Most importantly, trustworthiness is appraised even implicitly, indexing that it is a very relevant dimension. In Experiment 3 I demonstrated that perceived trustworthiness shapes VWM representations of faces even under conditions in which trustworthiness is task-irrelevant. This evidence may suggest that perceived trustworthiness may modulate in an automatic fashion a variety of psychological processes, including empathy.

Indeed, I in Experiments 4 and 5 I investigated the impact of implicitly perceived trustworthiness on empathy. It is well known that empathy may be modulated by social and affective relations between individuals. People tend to be more empathetic toward similar others − for instance in terms of group membership (see Bufalari & Ionta, 2013; Eres & Molenberghs, 2013) – as it has been shown, for instance in Experiment 2, with race. In this vein, neural responses associated with empathy are shaped by learned preferences and appraisal of others' social behaviour, such as social fairness. The issue addressed in Chapter 5 (Experiment 4 and 5) was whether empathy could be modulated even by the first impression people form of others, when information about social behaviour and personality traits is lacking.

I implemented a modified version of the pain decision task in which I presented participants with computerized (Experiment 4) and real (Experiment 5) trustworthy- and untrustworthy-looking faces associated with either a painful or neutral stimulation. The change in the pain decision task used in these two experiments consisted of the temporal separation between the face and the object evoking the painful (i.e. the needle of a syringe) or the neutral (i.e. the Q-tip) stimulation. This separation was essential to allow participants to implicitly appraise trustworthiness. Results of both Experiments 4 and 5 showed that perceived trustworthiness modulates empathy for pain. Trustworthy-looking faces induced in the observers magnified empathic responses relative to untrustworthy-looking faces, similarly to individuals known to be fair/unfair (Singer et al., 2006).

Interestingly, the naturalism of the stimuli affected empathy in a different fashion. Computerized faces induced magnified P3 reaction to pain, involving mentalizing mechanisms, whereas real faces induced magnified reaction to pain also in an earlier time-window pre-P3, namely in the N2-N3 time-window, involving experience sharing mechanisms. Consistently with these findings, Fan and Han (2008) demonstrated that cartoons relative to real stimuli elicit postponed empathic reaction to others' pain. In this vein, Zaki and Ochsner (2009) highlighted that realistic social contexts imply the dynamic integration of multimodal information that might shape brain activity in an unpredictable way while performing particular experimental tasks. Taken together, Fan and Han's study and the results of the Experiments 4 and 5 (i.e. with computerized and real faces, respectively) suggest that such integration involves experience sharing mechanisms that might need automaticity. Indeed, as Spunt and Lieberman (2011) suggest, naturalism of the stimuli likely acts as a connection, not an interaction, between mirror and mentalizing.

6.1 Is neural activity specific for empathy?

A critical aspect of studies on empathy towards others' pain is whether the reaction observed, at least the electrophysiological reaction, to pain is empathy-specific.

Noteworthy, that has never been demonstrated. In fact, empathy is a multicomponential process and cognitive sub-mechanisms might occur. That would qualify the reaction to pain as a general activity elicited in some affective context. Indeed, pain decision task procedure might involve overlapping mechanisms with those underlying an affective priming task where the interval between the face and the object is equal to zero as the case of Experiment 2, or in the context of the Experiments 4 and 5, is variable between 800 and 1600 ms. The affective priming is that effect occurring in a paradigm where two kinds of stimuli are presented sequentially and interleaved by a blank screen. The first stimulus (i.e., the prime) can either facilitate or inhibit participants' performance required on the second stimulus (i.e., the target). Behaviorally, faster reaction times and better accuracy in congruent trials (i.e., when the prime and the target have both either positive or negative valence) when compared with incongruent trials are classically observed. Although electrophysiological studies of affective priming have demonstrated that affective incongruency is reflected on larger N400 (e.g., Aguado, Dieguez-Risco, Méndez-Bértolo, Pozo, & Hinojosa, 2013; Eder, Leuthold, Rothermund, & Schweinberger, 2011) the findings of the current research might still be related to affective priming effect such as a negative shift of the the P300 for incongruent trials (e.g., untrustworthy faces with the Q-tip). That needs to be tested directly.

To rule out the possibility that previous findings of Experiment 4 and 5 were not specific to empathic reactions to pain, we administered some participants with an affective priming task. Previous findings have shown that the affective priming effect decays within 300 ms between the prime and target onsets (e.g., Fazio et al., 1986; Klauer & Musch, 2003). In line with this evidence, I conducted a control experiment in which the principal manipulation was the inter-stimulus interval (i.e., ISI) duration. I manipulated the ISI such that in one condition (i.e. short ISI) it was compatible with an affective priming paradigm and in in the other condition (i.e., long ISI) it was the exact mean of the variable ISI used for Experiments 4 and 5. I specifically focused on these last experiments to build this control because of the temporal separation between the face and the object that might have magnified possible underlying mechanisms of affective priming, less observable with pictures where the face was directly stimulated.

I would please the reader to see details of the sixth experiment, reported in Appendix 2.

Results of the Experiment 6 showed first of all no behavioral effects, so that I cannot directly conclude that the experimental manipulation did not succeed in eliciting an affective priming because of the fact that statistical non-significant effect is not informative per se. Most importantly, the main effect of the ISI was the only significant result of this experiment with larger centro-parietal P3 in long ISI condition (7.829 μV) relative to short ISI condition (4.506 μV). Most importantly there is no effect of trustworthiness, indicating that the mere presence of a face either trustworthy or untrustworthy does not affect the neural response on the object as evidence that no affective congruency is observed in the studies on empathy described in the present thesis.

6.2 So, what is empathy?

Edith Stein defined empathy in the early years of the 20th century as that ability of realizing: “empathy is a sui generis intentional state that reveals to us persons and their experiences”; she insisted on the immediacy of empathy as “read in another’s face”, further she distinguished it from “knowing another’s experience through inference or projecting one’s experience into them” (see Stein, 1964: 10).

This philosophic concept might be considered an early line of reasoning about what at present is known as *simulation theory* (ST), which is encompassed in the Theory of Mind[17]. As opposed to *theory theory* (TT), ST states that people understand others’ mental states by taking the other perspective and using one’s own experience and resources to simulate the other’s mind. According to ST one can understand the other’s mind because it is *as if* the other would be the self. Goldman and Jordan (2013), endorsing ST, proposed a bi-level model of mindreading according to which low- and high-level simulational processes are distinguished. Whereas low-level simulational processes trace mirror neuron system account, high-level processes encompass imagination.

Goldman and Jordan (2013) suggest that through imagination people recreate the feeling of a state and hold it in mind to simulate it, to feel like real first-hand

---

[17] The theory according to which people attribute mental states to oneself and to the others and understand that the others might have mental states different from one’s own. The Theory of Mind (ToM) encompasses two main accounts: The Theory Theory (TT), which is the predominant theory and states that our understanding of others’ mind is possible through the development of a “commonsens” along the lifetime that allow people to predict others’ behavior; and the Simulation Theory (ST).

experience, and so understanding the other’s perspective. Imagining is a real-like experience through which simulation is possible. According to the authors, indeed, mindreading needs “shared representation” and imagination is the gate.

In the light of the studies described in the present thesis, what I think of empathy might be endorsed in Goldman and Jordan’s view. Empathy is indeed, first of all, that specific ability that allows people to use inner experience to grasp, with large degree of certainty, other’s inner state and so allowing to react properly in the emerging interaction.

Results of Experiment 1 might be refined in the light of Goldman and Jordan’s account. The sensorimotor cue triggered simulation *per se*, i.e. mechanisms more automatic of neural resonance, the contextual cue triggered simulation through imagination. However, the presence of both cues did not magnify empathic reaction, suggesting that people process the available information but exploit only the necessary to react properly, quickly and parsimoniously. Further there is an intermediate time-window in which unspecific response that might act as a transfer from experience sharing to mentalizing is observable. An alternative view might proposes that intermediate time-window as temporal overlapping between the two subprocesses that can share mechanisms due to the specific context or task requirements. This needs to be further investigated by future research.

So experience sharing and mentalizing are functionally dissociable through the selective engagement of one or the other in response to specific cues.

Computerized faces used in Experiment 4 and the specific experimental design totally suppressed the empathic response in the experience sharing time-window, whereas the presence of more ecologically valid stimuli, as those used in Experiment 2 and 5, restored such response.

On the basis of the results of the presented studies, I conclude that the more people need to imagine the other's state, the more mentalizing is selectively engaged; the more visual information is available, the lesser people need to imagine the other's state and concurrently the more neural resonance is involved that in turns activates experience sharing (Figure 1).

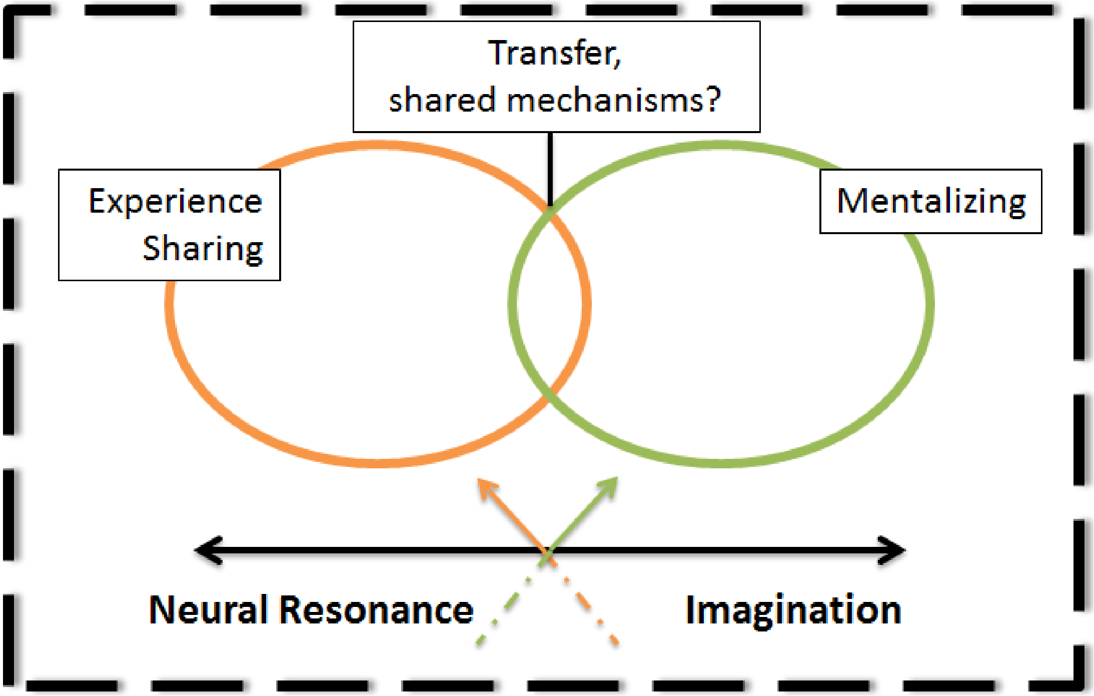


Figure 1. Schematic representation of the proposed model of empathy.

Appendix 1
EQ. *Italian version.*

Istruzioni

1. Inserire nome (facoltativo).
2. Inserire genere.
3. Leggere attentamente le 60 affermazioni e marcare la casella inerente alla risposta prescelta (non piu' di una ad affermazione).

Nome e cognome:
○ Uomo ○ Donna

	ASSOLUTAMENTE D'ACCORDO	PARZIALMENTE D'ACCORDO	LEGGERMENTE IN DISACCORDO	ASSOLUTAMENTE CONTRARIO
1. Capisco con facilità se qualcuno vuole partecipare ad una conversazione.				
2. Preferisco gli animali agli esseri umani.				
3. Provo a seguire le nuove mode e le nuove tendenze.				
4. Quando un concetto per me facilmente intuibile non viene compreso alla prima spiegazione, ho difficoltà a rispiegarlo.				
5. Sogno la maggior parte delle notti.				
6. Prendermi cura degli altri è qualcosa che mi fa veramente piacere.				

	ASSOLUTAMENTE D’ACCORDO	PARZIALMENT E D’ACCORDO	LEGGERMENTE IN DISACCORDO	ASSOLUTAMENTE CONTRARIO
7. Provo a risolvere da solo i miei problemi piuttosto che discuterne con gli altri.				
8. Trovo difficile capire come comportarmi in mezzo alla gente.				
9. Sono al massimo della mia forma nelle prime ore della giornata.				
10. La gente mi dice spesso che insisto troppo sui miei argomenti.				
11. Non mi preoccupa più di tanto essere in ritardo ad un appuntamento con un amico.				
12. Mi tengo lontano da amicizie e relazioni sociali, dal momento che ritengo siano troppo difficili da curare.				
13. Non infrangerei mai una legge, seppur minima.				
14. Spesso ho difficoltà nel distinguere le buone dalle cattive maniere.				
15. In una conversazione tendo ad incentrare il mio discorso sul mio modo di pensare piuttosto che su quello degli altri.				
16. Preferisco gli scherzi all’ ironia.				
17. Preferisco vivere il presente piuttosto che pensare al futuro.				
18. Quando ero bambino/a mi divertivo a sezionare i vermi per vedere cosa succedeva.				
19. Riesco facilmente a capire se qualcuno dice una cosa ma ne intende un’ altra				

	ASSOLUTAMENTE D'ACCORDO	PARZIALMENTE D'ACCORDO	LEGGERMENTE IN DISACCORDO	ASSOLUTAMENTE CONTRARIO
20. Sono molto moralista				
21. Non capisco perché la gente si offende tanto per certe cose.				
22. Riesco facilmente a mettermi nei panni degli altri.				
23. Penso che le buone maniere siano la cosa più importante che un genitore possa insegnare al proprio figlio.				
24. Mi piace agire d'istinto.				
25. Sono bravo/a a prevedere i sentimenti degli altri.				
26. Mi accorgo subito se qualcuno in un gruppo è a disagio o imbarazzato.				
27. Se ciò che dico offende qualcuno, penso che non sia un mio problema ma di chi si sente offeso.				
28. Se qualcuno mi chiede un parere sul suo nuovo taglio di capelli rispondo sinceramente anche se non mi piace.				
29. Non riesco sempre a capire perché qualcuno potrebbe sentirsi offeso da certe affermazioni.				
30. La gente mi dice spesso che sono totalmente imprevedibile.				
31. Amo stare al centro delle attenzioni nelle situazioni di gruppo.				
32. Vedere qualcuno piangere non mi turba più di tanto.				

	ASSOLUTAMENTE D'ACCORDO	PARZIALMENTE D'ACCORDO	LEGGERMENTE IN DISACCORDO	ASSOLUTAMENTE CONTRARIO
33. Mi piace discutere di				
34. Il mio essere diretto viene spesso interpretato come scortesia anche se non è questa la mia intenzione.				
35. Non mi confondo nelle situazioni formali.				
36. La gente mi dice che sono bravo/a a capire ciò che sente e pensa.				
37. Quando parlo con la gente tendo più a discutere delle loro esperienze che delle mie.				
38. Mi turba veder soffrire un animale.				
39.Riesco a prendere le mie decisioni senza lasciarmi influenzare dai sentimenti degli altri.				
40. Non riesco a rilassarmi fin quando non concludo tutto ciò che ho pianificato per quel giorno.				
41. Riesco facilmente a capire se ciò che dico annoia o interessa qualcuno.				
42. Mi turbano le immagini di gente che soffre quando guardo le notizie in tv.				
43. Gli amici spesso si confidano con me perché dicono che capisco bene i loro problemi				
44. Riesco a percepire se la mia presenza è indesiderata, anche se non mi viene detto espressamente.				
45. Mi creo spesso nuovi hobbies ma mi annoio facilmente e passo ad altro				

	ASSOLUTAMENTE D’ACCORDO	PARZIALMENT E D’ACCORDO	LEGGERMENTE IN DISACCORDO	ASSOLUTAMENTE CONTRARIO
46. Talvolta la gente mi dice che esagero nello scherzo.				
47. Mi innervosirebbe troppo fare un giro su grandi montagne russe.				
48. La gente mi dice spesso che sono insensibile ma io non capisco il perché.				
49. Se vedo un estraneo in un gruppo penso che stia a lui fare uno sforzo per inserirsi.				
50. Solitamente i film non mi coinvolgono emotivamente.				
51. Mi piace essere perfettamente organizzato/a nella vita di tutti i giorni e spesso faccio una lista delle				
52. Riesco a percepire in modo rapido e intuitivo come qualcun altro si sente.				
53. Non mi piace rischiare				
54. Riesco facilmente ad intuire ciò di cui un’ altra persona vorrebbe parlare.				
55. Capisco se qualcuno sta celando le sue emozioni.				
56. Prima di prendere una decisione valuto tutti i pro e i contro.				
57. Non rifletto sulle regole da rispettare quando sono tra la gente.				
58. Sono bravo/a a prevedere le mosse degli altri.				
59. I problemi degli amici mi coinvolgono emotivamente.				
60. Di solito tengo in considerazione i punti di vista degli altri anche se non li condivido.				

EQ. *English version.*

Istructions

1. Please fill in the name (optional).
2. Please fill in the sex .
3. Please then read each statement <u>very carefully</u> and rate how strongly you agree or disagree with it by circling your answer. There are no right or wrong answers, or trick questions. IN ORDER FOR THE SCALE TO BE VALID, YOU MUST ANSWER EVERY QUESTION.

Name:
○ Male ○ Female

	STRONGLY AGREE	SLIGHTLY AGREE	SLIGHTLY DISAGREE	STRONGLY DISAGREE
1. I can easily tell if someone else wants to enter a conversation.				
2. I prefer animals to humans.				
3. I try to keep up with the current trends and fashions.				
4. I find it difficult to explain to others things that I understand easily, when they don’t understand it first time.				
5. I dream most nights.				
6. I really enjoy caring for the other people.				

	STRONGLY AGREE	SLIGHTLY AGREE	SLIGHTLY DISAGREE	STRONGLY DISAGREE
7. I try to solve my own problems rather than discussing them with others.				
8. I find it hard to know what to do in a social situation.				
9. I am at my best first thing in the morning.				
10. People often tell me that I went too far in driving my point home in a discussion.				
11. It doesn’t bother me too much if I am late meeting a friend.				
12. Friendships and relationships are just too difficult, so I tend not to bother with them.				
13. I would never break a law, no matter how minor.				
14. I often find it difficult to judge if something is rude or polite.				
15. In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.				
16. I prefer practical jokes to verbal humor.				
17. I live for today rather than the future.				
18. When I was a child, I enjoyed cutting up worms to see what would happen.				
19. I can pick up quickly if someone says one thing but means another.				

	STRONGLY AGREE	SLIGHTLY AGREE	SLIGHTLY DISAGREE	STRONGLY DISAGREE
20. I tend to have very strong opinions about morality.				
21. It is hard for me to see why some things upset people so much.				
22. I find it easy to put myself in somebody else's shoes.				
23. I think that good manners are the most important thing a parent can teach their child.				
24. I like to do things on the spur of the moment.				
25. I am good at predicting how someone will feel.				
26. I am quick to spot when someone in a group is feeling awkward or uncomfortable.				
27. If I say something that someone else is offended by, I think that that's their problem, not mine.				
28. If anyone asked me if I liked their haircut, I would reply truthfully, even if I didn't like it.				
29. I can't always see why someone should have felt offended by a remark.				
30. People often tell me that I am very unpredictable.				
31. I enjoy being the centre of attention at any social gathering.				
32.Seeing people cry doesn't reall upset me.				

	STRONGLY AGREE	SLIGHTLY AGREE	SLIGHTLY DISAGREE	STRONGLY DISAGREE
33. I enjoy having discussions about politics.				
34. I am very blunt, which some people take to be rudeness, even though this is unintentional.				
35. I don’t tend to find social situations confusing.				
36. Other people tell me I am food at understanding how they are feeling and what they are thinking.				
37. When I talk to people, I tend to talk about their experiences rather than my own.				
38. It upsets me to see an animal in pain.				
39.I am able to make decisions without being influenced by people’s feelings.				
40. I can’t relax until I have done everything I had planned to do that day.				
41. I can easily tell if someone else is interested or bored with what I am saying.				
42. I get upset if I see people suffering on news programs.				
43. Friends usually talk to me about their problems as they say that I am very understanding.				
44. I can sense if I am intruding, even if the other person doesn’t tell me.				
45. Mi creo spesso nuovi hobbies ma mi annoio facilmente e passo ad altro				

	STRONGLY AGREE	SLIGHTLY AGREE	SLIGHTLY DISAGREE	STRONGLY DISAGREE
46. People sometimes tell me that I have gone too far with teasing.				
47. I would be too nervous to go on a big rollercoaster.				
48. Other people often say that I am insensitive, though I don’t always see why.				
49. If I see a stranger in a group ,I think that it is up to them to make an effort to join in.				
50. I usually stay emotionally detached when watching a film.				
51. I like to be very organised in day to day life and often make lists of the chores I have to do.				
52. I can tune into how someone else feels rapidly and intuitively.				
53. I don’t like to take risks.				
54. I can easily work out what another person might want to talk about.				
55. I can tell if someone is masking their true emotion.				
56. Before making a decision I always weigh up the pros and cons.				
57. I don’t consciously work out the rules of social situations.				
58. I am good at predicting what someone will do.				
59. I tend to get emotionally involved with a friend’s problems.				
60. I can usually appreciate the other person’s viewpoint, even if I don’t agree with it.				

Interpersonal Reactivity Index. *Italian version.*

Troverai ora una lista di affermazioni che possono essere più o meno vere / false per te.

Non ci sono risposte giuste o sbagliate: basati sulle tue sensazioni ed opinioni.

Ti chiediamo di leggere attentamente ciascuna affermazione e di indicare la tua opinione con i numeri:

- **1** se essa è **Mai vera** per te
- **2** se essa è **Raramente vera** per te
- **3** se essa è **Qualche volta vera** per te
- **4** se essa è **Spesso vera** per te
- **5** se essa è **Sempre vera** per te

1 **Mai vera per me**	**2** **Raramente vera per me**	**3** **Qualche volta vera per me**	**4** **Spesso vera per me**	**5** **Sempre vera per me**

1. Sogno ad occhi aperti e fantastico, con una certa regolarità, sulle cose che potrebbero accadermi.	1	2	3	4	5
2. Provo spesso sentimenti di tenerezza e di preoccupazione per le persone meno fortunate di me.	1	2	3	4	5
3. A volte trovo difficile vedere le cose dal punto di vista di un’altra persona.	1	2	3	4	5
4. A volte non mi sento particolarmente dispiaciuto/a per le altre persone che hanno problemi.	1	2	3	4	5
5. Resto veramente coinvolto/a dagli stati d’animo dei protagonisti di un racconto.	1	2	3	4	5
6. In situazioni d’emergenza, mi sento apprensivo e a disagio.	1	2	3	4	5
7. Riesco solitamente ad essere obiettivo/a quando guardo un film o una rappresentazione teatrale e raramente mi lascio coinvolgere del tutto.	1	2	3	4	5
8. In caso di disaccordo, cerco di tenere conto del punto di vista di ognuno prima di prendere una decisione.	1	2	3	4	5
9. Quando vedo qualcuno che viene sfruttato, provo sentimenti di protezione nei suoi confronti.	1	2	3	4	5

10. A volte mi sento indifeso/a quando mi trovo in situazioni emotivamente coinvolgenti.	1	2	3	4	5
11. Qualche volta cerco di comprendere meglio i miei amici immaginando come appaiono le cose dalla loro prospettiva.	1	2	3	4	5
12. Mi accade raramente di essere coinvolto/a da un buon libro o da un bel film.	1	2	3	4	5
13. Quando vedo qualcuno farsi male tendo a rimanere calmo.	1	2	3	4	5
14. Le disgrazie degli altri solitamente non mi turbano molto.	1	2	3	4	5
15. Se sono sicuro di avere ragione su qualcosa, non perdo tempo ad ascoltare le ragioni degli altri.	1	2	3	4	5
16. Dopo aver visto una commedia o un film mi sento come se fossi stato uno dei protagonisti.	1	2	3	4	5
17. Mi spaventa il fatto di trovarmi in situazioni che provocano tensione emotiva.	1	2	3	4	5
18. Quando vedo qualcuno che viene trattato ingiustamente, talvolta mi capita di non provare molta pietà per lui.	1	2	3	4	5
19. Solitamente sono molto efficace nel far fronte alle situazioni d’emergenza.	1	2	3	4	5
20. Spesso mi sento abbastanza colpito dalle cose che vedo accadere.	1	2	3	4	5
21. Credo che ci siano due prospettive diverse per ogni questione e cerco di capirle entrambe.	1	2	3	4	5
22. Mi descriverei come una persona dal cuore piuttosto tenero.	1	2	3	4	5
23. Quando guardo un bel film riesco facilmente ad immedesimarmi nel personaggio principale.	1	2	3	4	5
24. Tendo a perdere il controllo durante le emergenze.	1	2	3	4	5
25. Quando sono in contrasto con qualcuno, solitamente provo a “mettermi nei suoi panni” per un po’.	1	2	3	4	5
26. Quando leggo una storia o un romanzo interessante, immagino come mi sentirei se gli avvenimenti della storia accadessero a me.	1	2	3	4	5
27. Quando vedo qualcuno che in una situazione di emergenza necessita disperatamente di aiuto, vado in crisi.	1	2	3	4	5
28. Prima di criticare qualcuno provo ad immaginare come mi sentirei se fossi al suo posto.	1	2	3	4	5

Interpersonal Reactivity Index. *Engligh version.*

The following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by choosing the appropriate letter on the scale at the top of the page: A, B, C, D, or E. When you have decided on your answer, fill in the letter on the answer sheet next to the item number. READ EACH ITEM CAREFULLY BEFORE RESPONDING. Answer as honestly as you can. Thank you.

ANSWER SCALE:

A **B** **C** **D** **E**

DOES NOT DESCRIBE ME WELL DESCRIBES ME VERY WELL

1. I daydream and fantasize, with some regularity, about things that might happen to me.	A E	B	C	D
2. I often have tender, concerned feelings for people less fortunate than me.	A E	B	C	D
3. I sometimes find it difficult to see things from the "other guy's" point of view.	A E	B	C	D
4. Sometimes I don't feel very sorry for other people when they are having problems.	A E	B	C	D
5. I really get involved with the feelings of the characters in a novel.	A E	B	C	D
6. In emergency situations, I feel apprehensive and ill-at-ease.	A E	B	C	D
7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.	A E	B	C	D
8. I try to look at everybody's side of a disagreement before I make a decision.	A E	B	C	D
9. When I see someone being taken advantage of, I feel kind of protective towards them.	A E	B	C	D
10. I sometimes feel helpless when I am in the middle of a very emotional situation.	A E	B	C	D
11. I sometimes try to understand my friends better by imagining how things look from their perspective.	A E	B	C	D
12. Becoming extremely involved in a good book or movie is somewhat rare for me.	A E	B	C	D
13. When I see someone get hurt, I tend to remain calm.	A E	B	C	D
14. Other people's misfortunes do not usually disturb me a great deal.	A E	B	C	D

15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.	A E	B	C	D
16. After seeing a play or movie, I have felt as though I were one of the characters.	A E	B	C	D
17. Being in a tense emotional situation scares me.	A E	B	C	D
18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.	A E	B	C	D
19. I am usually pretty effective in dealing with emergencies.	A E	B	C	D
20. I am often quite touched by things that I see happen.	A E	B	C	D
21. I believe that there are two sides to every question and try to look at them both.	A E	B	C	D
22. I would describe myself as a pretty soft-hearted person.	A E	B	C	D
23. When I watch a good movie, I can very easily put myself in the place of a leading character.	A E	B	C	D
24. I tend to lose control during emergencies.	A E	B	C	D
25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.	A E	B	C	D
26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.	A E	B	C	D
27. When I see someone who badly needs help in an emergency, I go to pieces.	A E	B	C	D
28. Before criticizing somebody, I try to imagine how I would feel if I were in their place.	A E	B	C	D

Appendix 2

Experiment 6

Noteworthy, it has never been demonstrated that electrophysiological reaction to pain is empathy-specific. In fact, empathy is a multicomponential process and cognitive sub-mechanisms might occur. That would qualify the reaction to pain as a general activity elicited in some affective context. Indeed, pain decision task procedure might have overlapping mechanisms with those underlying an affective priming task where the interval between the face and the object is equal to zero, or in the context of the current studies, is variable between 800 and 1600 ms. The affective priming is that effect occurring in a paradigm where two kinds of stimuli are presented sequentially and interleaved by a blank screen. The first stimulus (i.e., the prime) can either facilitate or inhibit participants' performance required on the second stimulus (i.e., the target). Behaviorally, faster reaction times and better accuracy in congruent trials (i.e., when the prime and the target have both either positive or negative valence) when compared with incongruent trials are classically observed. Although electrophysiological studies of affective priming have demonstrated that affective incongruency is reflected on larger N400 (e.g., Aguado, Dieguez-Risco, Méndez-Bértolo, Pozo, & Hinojosa, 2013; Eder, Leuthold, Rothermund, & Schweinberger, 2011) findings of the current research might still be related to affective priming effect such as a negative shift of the P300 for incongruent trials (e.g., untrustworthy faces with the Q-tip) and it needs to be tested directly.

To rule out the possibility that previous findings were not specific to empathic reactions to pain, I administered some participants with an affective priming task.

Previous findings have shown that the affective priming effect decays within 300 ms between the prime and target onsets (e.g., Fazio et al., 1986; Klauer & Musch, 2003). In line with this evidence, the principal manipulation in Experiment 6 was the inter-stimulus interval (i.e., ISI) duration. We manipulated the ISI such that in one condition (i.e. short ISI) it was compatible with an affective priming paradigm and in in the other condition (i.e., long ISI) it was the exact mean of the variable ISI used for Experiment 4 and Experiment 5.

Stimuli

The stimuli used are the same as those used for Experiment 4.

Affective priming task. Participants performed an affective priming task, which has been implemented similarly to the pain decision task. Figure 1 shows a schematic illustration of the paradigm.

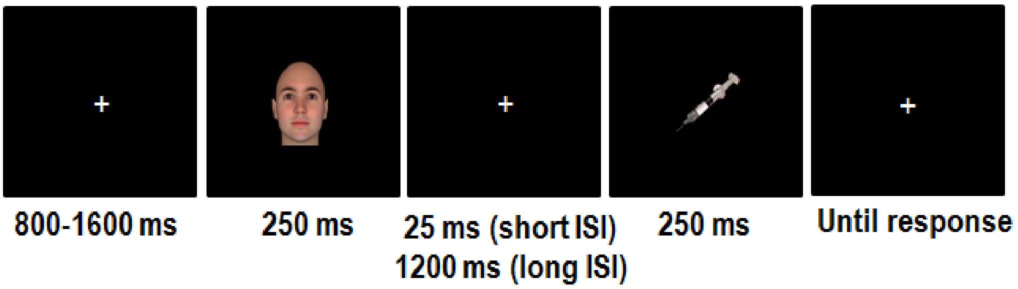


Figure 1. Schematic illustration of the paradigm, adapted to be an affective priming by manipulating the interval between the face (i.e., here the prime) and the object (i.e., here the target).

Each trial began with a randomly variable fixation cross at the center of the computer screen (800-1600 ms in steps of 100 ms). A trustworthy or untrustworthy face (presented for 200 ms) and either a Q-tip or a syringe (presented for 250 ms) evocating a neutral or a painful stimulation respectively, were interleaved by a blank screen presented for either 25 ms (short ISI) or 1200 ms (long ISI). Notably, the long ISI falls exactly in the mean of the variable interval between the face and the object implemented in the pain decision task (800-1600 ms in steps of 100 ms). Participants were instructed to observe the face but to decide whether the stimulation evocated by the object was either painful or neutral by pressing one of two response keys (“F” and “J”) of the computer keyboard. In the 20% of the trials the object was replaced by another face, which could be either identical or different to the first one. In those cases participants were required to say whether the faces were identical or not.

The different types of trials were randomly intermixed. Participants were instructed to perform the task as faster and accurately as possible.

EEG acquisition and analysis. The EEG recording and analysis was the same as in Experiment 1 and Experiment 2. The time window considered for statistical analysis in this experiment was of 310-820 ms.

Priming task. Individual mean proportions of correct responses were submitted to a repeated measure analysis of variance (ANOVA) considering stimulation condition (painful vs. non-painful) and trustworthiness (trustworthy faces vs. untrustworthy faces) as within-subjects factors. ANOVA did not show significant factor effects ($F < 1$).

P3. Figure 2 shows P3 waveforms time-locked to syringe/Q-tip elicited at fronto-central and centro-parietal sites in painful and non-painful conditions and separately for trustworthy and untrustworthy faces and for short and long ISI. Statistical analyses on the fronto-central and centro-parietal P3 mean amplitude values were submitted to a 2 (stimulation: painful vs. non-painful) x 2 (trustworthiness: trustworthy vs. untrustworthy faces) x 2 (ISI: short vs. long) repeated measures ANOVA as within-subjects factors. The centro-parietal P3 was larger in long ISI (7.829 μV) relative to short ISI condition (4.506 μV) as the main effect of ISI was significant at centro parietal sites, $F(1, 9) = 16.727$, $p < .001$; $\eta_p^2 > .650$).

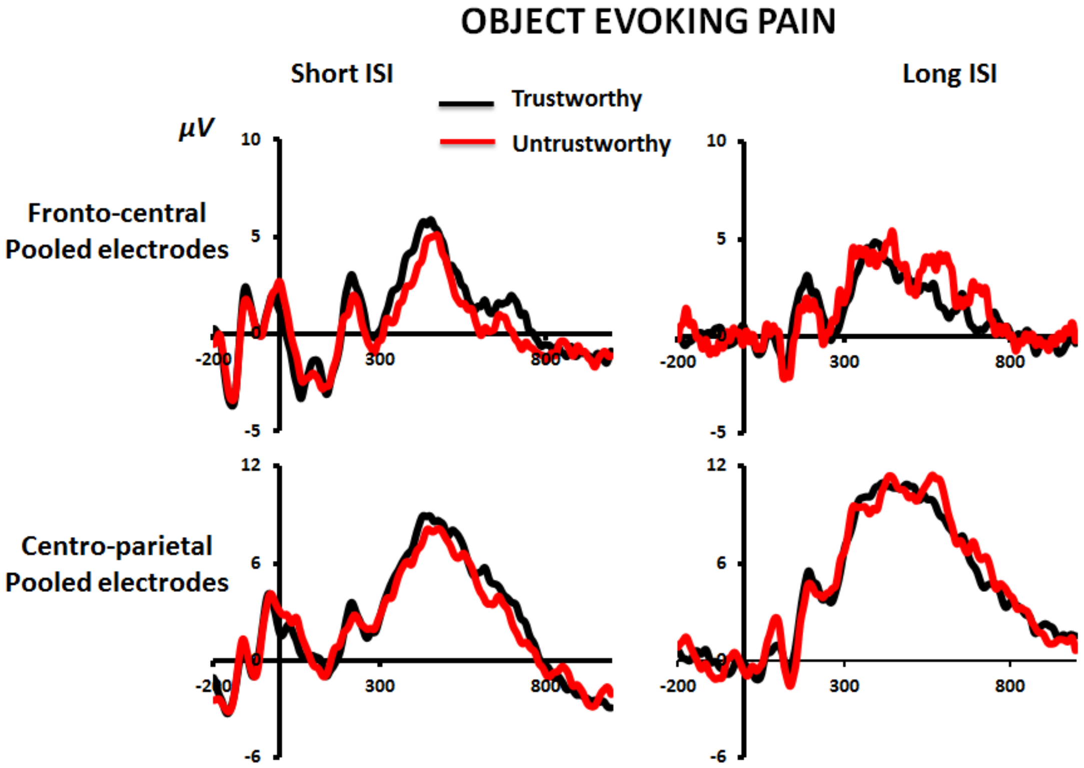


Figure 2. P3 components in response to the syringe (i.e., the object evoking pain) for trustworthy and untrustworthy faces. This is the most important comparison because it shows how trustworthiness, per se, does not prime a congruent/incongruent affective response to the object that should trigger an empathic response to pain.

Most importantly, trustworthiness nor other main effects or interaction between the factors approached significance (all *Fs* <1 at fronto-central sites; max *F*=3.018 min *p*=.116 at centro-parietal sites).

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews,* 1, 21–62.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, 393, 470–474.

Aguado, L., Dieguez-Risco, T., Méndez-Bértolo, C., Pozo, M. A., & Hinojosa, J. A. (2013). Priming effects on the N400 in the affective priming paradigm with facial expressions of emotion. *Cognitive, Affective, & Behavioral Neuroscience*, 1-13.

Albiero, P., Ingoglia, S., & Lo Coco, A. (2006). Albiero, P., Ingoglia, S., Lo Coco, A. (2006). Contributo all'adattamento italiano dell'Interpersonal Reactivity Index [A contribution to the Italian validation of the Interpersonal Reactivity Index]. *Testing Psicometria Metodologia*, 13, 107–125.

Amodio, D. M. (2010). Coordinated Roles of Motivation and Perception in the Regulation of Intergroup Responses : Frontal Cortical Asymmetry Effects on the P2 Event-related Potential and Behavior. *Journal of Cognitive Neuroscience*, 22, 2609–2617.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology,* 91, 652-661.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience,* 7, 268–77.

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology*, 84, 738–753.

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, 20, 143-148.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011a). Learned value magnifies salience-based attentional capture. *PloS One*, 6, e27926.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011b). Value-driven attentional capture. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 10367–10371.

Asendorpf, J., & Baudonnière, P. (1993). Self-awareness and other-awareness: Mirror self-recognition and synchronic imitation among unfamiliar peers. *Developmental Psychology*, 29, 88–95.

Attwood, A. S., Penton-Voak, I. S., Burton, A. M., & Munafò, M. R. (2013). Acute Anxiety Impairs Accuracy in Identifying Photographed Faces. *Psychological Science*, 24, 1591–1594.

Awh, E., Barton, B., Vogel, E.K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18, 622–628.

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience*, 8, 955–960.

Avenanti, A., Minio-Paluello, I., Sforza, A., & Aglioti, S. M. (2009). Freezing or escaping? Opposite modulations of empathic reactivity to the pain of others. *Cortex*, 45, 1072–7.

Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology*, 20, 1018-1022.

Azevedo, R. T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S. M. (2012). Their pain is not our pain: brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping*. 34, 3168–81.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, 6, 269.

Baron, S. G., Gobbini, M. I., Engell, A. D., & Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social Cognitive and Affective Neuroscience*, 6, 572–581.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163-175.

Benuzzi, F., Lui, F., Duzzi, D., Nichelli, P. F., & Porro, C. A. (2008). Does it look painful or disgusting? Ask your parietal and cingulate cortex. *The Journal of Neuroscience*, 28, 923–31.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain*. 133, 1515–25.

Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, Alvaro Brugger, P., Seeck, M., Landis, T., & Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *The Journal of neuroscience*, 25, 550–557.

Boone, R., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, 27, 163–182.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. a, Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage*, 25, 312–9.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes versus action simulation. *Current Biology*, 17, 2117–2121.

Bufalari, I., & Ionta, S. (2013). The social and personality neuroscience of empathy for pain and touch. *Frontiers in Human Neuroscience*, 7, 393.

Bzdok, D., Langner, R., Caspers, S., Kurth, F., Habel, U., Zilles, K., Laird, A., & Eickhoff, S. B. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure & Function*, 215, 209–223.

Calvo, M. G., Eysenck, M. W., & Castillo, M. D. (1997). Interpretation bias in test anxiety: The time course of predictive inferences. *Cognition & Emotion*, 11, 43–64.

Canadian Medical Association Journal (2011, January 24). Showing empathy to patients can improve care. *ScienceDail*y. Retrieved June 23, 2013, from http://www.sciencedaily.com- /releases/2011/01/110124121543.htm.

Carpenter, M., Uebel, J., & Tomasello, M. (2013). Being mimicked increases prosocial behavior in 18-month-old infants. *Child Development*, 84, 1511–8.

Chakrabarti, B., Bullmore, E., Baron-Cohen, S. (2006). Empathizing with basic emotions: common and discrete neural substrates. *Social Neuroscience*, 1, 364–84.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, 25, 215–224.

Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D., & Decety, J. (2007). Expertise modulates the perception of pain in others. *Current Biology*, 17, 1708–1713.

Cheng, Y., Yang, C. Y., Lin, C. P., Lee, P. L., & Decety, J. (2008). The perception of pain in others suppresses somatosensory oscillations: a magnetoencephalography study. *Neuroimage*, 40, 1833-1840.

Cheon, B. K., Im, D. M., Harada, T., Kim, J. S., Mathur, V. A., Scimeca, J. M., ... & Chiao, J. Y. (2011). Cultural influences on neural basis of intergroup empathy. *Neuroimage*, 57, 642-650.

Chiao, J. Y., & Mathur, V. A. (2010). Intergroup empathy: how does race affect empathic neural responses?. *Current Biology*, 20, R478-R480.

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of personality and social psychology*, 73, 481.

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, 30, 203–216.

Cochin, S., Barthelemy, C., Roux, S., & Martineau, J. (1999). Observation and execution of movement: similarities demonstrated by quantified electroencephalography. *European Journal of Neuroscience*, 11, 1839-1842.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of Emotions*, 2, 91–115.

Craig, A. D. (2003a). A new view of pain as a homeostatic emotion. *Trends in Neurosciences*, 26, 303-307.

Craig, A. D. (2003b). Pain mechanisms: labeled lines versus convergence in central processing. *Annual review of neuroscience*, 26, 1-30.

Craig, A. D. (2009). How do you feel—now? the anterior insula and human awareness.

Cunningham, W. a, Johnson, M. K., Raye, C. L., Chris Gatenby, J., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, 15, 806–13.

Danziger, N., Prkachin, K. M., & Willer, J. C. (2006). Is pain the price of empathy? The perception of others' pain in patients with congenital insensitivity to pain. *Brain*, 129, 2494-2507.

David, N., Aumann, C., Santos, N. S., Bewernick, B. H., Eickhoff, S. B., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2008). Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. *Social Cognitive and Affective Neuroscience*, 3, 279–289.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology,* 44, 113–126.

De Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 16701–6.

de Greck, M., Wang, G., Yang, X., Wang, X., Northoff, G., & Han, S. (2012). Neural substrates underlying intentional empathy. *Social cognitive and affective neuroscience*, 7, 135–44.

De Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, 10, 435–41.

De Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual review of psychology*, 59, 279–300.

Decety, J. (2011). Dissecting the Neural Mechanisms Mediating Empathy. *Emotion Review*, 3, 92–108.

Decety, J, Echols, S., & Correll, J. (2009). The Blame Game : The Effect of Responsibility and Social Stigma on Empathy for Pain. *Journal of Cognitive Neuroscience*, 22, 985–997.

Decety, J., & Jackson, P. (2006). A social-neuroscience perspective on empathy. *Current Directions in Psychological Science,* 15, 54–58.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews,* 3, 71–100.

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *Scientific World Journal*, 6, 1146–163.

Decety, J, Yang, C. Y., & Cheng, Y. (2010). Physicians down-regulate their pain empathy response: an event-related brain potential study. *Neuroimage,* 50, 1676–82.

Dell'Acqua, R., Pesciarelli, F., Jolicour, P., Eimer, M., & Peressotti, F. (2007). The interdependence of spatial attention and lexical access as revealed by early asymmetries in occipito-parietal ERP activity. *Psychophysiology,* 44, 436-443.

Dell'Acqua, R., Sessa, P., Toffanin, P., Luria, R., & Joliccoeur, P. (2010). Orienting attention to objects in visual short-term memory. *Neuropsychologia,* 48, 419-428.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research,* 91, 176–180.

Dickter, C. L., & Bartholow, B. D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, 2, 189–98.

Donchin, E. (1981). Surprise! . . .. Surprise? *Psychophysiology*, 18, 493–515.

Dum, R. P., Levinthal, D. J., & Strick, P. L. (2009). The spinothalamic system targets motor and sensory areas in the cerebral cortex of monkeys. *The Journal of Neuroscience*, 29, 14223–14235.

Dzhelyova, M., Perrett, D. I., & Jentzsch, I. (2012). Temporal dynamics of trustworthiness perception. *Brain Research*, 1435, 81–90.

Eberhardt, J. L. (2002). Imaging race. *The American psychologist*, 60, 181–90.

Edele, A., Dziobek, I., & Keller, M. (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and Individual Differences*, 24, 96–102.

Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2012). Automatic response activation in sequential affective priming: an ERP study. *Social Cognitive and Affective Neuroscience*, 7, 436-445.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, 99, 225-234.

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, 1508–1519.

Eres, R., & Molenberghs, P. (2013). The influence of group membership on the neural correlates involved in empathy. *Frontiers in Human Neuroscience*, 7, 176.

Fan, Y., & Han, S. (2008). Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. *Neuropsychologia*, 46, 160–73.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.

Forgiarini, M., Gallucci, M., & Maravita, A. (2011). Racism and the empathy for pain on our skin. *Frontiers in Psychology*, 2, 108.

Foroni, F., & Semin, G. R. (2011). When does mimicry affect evaluative judgment? *Emotion*, 11, 687–690.

Frith, C., & Frith, U. (1999). Interacting minds: A biological basis. Science, 286, 1692–1695.

Frith, C., & Frith, U. (2006). The neural basis of mentalizing. *Neuron,* 50, 531–4.

Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 358, 459–73.

Frith, U., & Frith, C. (2001). The biological basis of social interaction. *Current Directions in Psychological Science*, 10, 151–155.

Frith, C., & Singer, T. (2008). The role of social cognition in decision making The role of social cognition in decision making. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363, 3875–3886.

Frühholz, S., Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *Neuroimage*, 62, 1658–66.

Gable, S. L., Reis, H. T., & Downey, G. (2003). He Said, She Said A Quasi-Signal Detection Analysis of Daily Interactions Between Close Relationship Partners. *Psychological Science*, 14, 100-105.

Gallese, V. (2001). The “Shared Manifold” Hypothesis. Journal of Consciousness Studies, 8, 33–50.

Gallese, V. (2003). The Roots of Empathy: The Shared Manifold Hypothesis and the Neural Basis of Intersubjectivity. *Psychopathology,* 36, 171–180.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396–403.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593–609.

Gao, T., Scholl, B. J., & McCarthy, G. (2012). Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *The Journal of neuroscience*, 32, 14276–80.

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current biology,* 16, 1824-1829.

Gobbini, M., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage*, 22, 1628–1635.

Goldman, A. I., & Jordan, L. C. (2013). Mindreading by simulation: The roles of imagination and mirroring. *Understanding Other Minds: Perspectives from developmental social neuroscience,* 448.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger.

Greenwald, A. G., Nosek, B. a., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197–216.

Gu, X., & Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *Neuroimage*, 36, 256–67.

Gutsell, J. N., & Inzlicht, M. (2010). Empathy constrained: Prejudice predicts reduced mental simulation of actions during observation of outgroups. *Journal of Experimental Social Psychology*, 46, 841–845.

Haberlandt, K. (1997). Cognitive psychology (2nd ed.). Boston: Allyn & Bacon.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. *Psychological Science*, 17, 847–53.

Hart, A. J., Whalen, P. J., Shin, L. M., Mcinerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*, 11, 2351–2355.

Heil, M., Rolke, B., & Pecchinenda, A. (2004). Automatic semantic activation is no myth: Semantic context effects on the N400 in the letter-search task in the absence of response time effects. *Psychological Science*, 15, 852–857.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, 68, 149–60.

Hein, G., & Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Current opinion in neurobiology*, 18, 153–8.

Hickok, G. (2008). Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. *Journal of Cognitive Neuroscience*, 21, 1229–1243.

Hirsch, C. R., & Clark, D. M. (2004). Information-processing bias in social phobia. *Clinical Psychology Review*, 24, 799–825.

Hofmann, S. G., Ellard, K. K., & Siegle, G. J. (2012). Neurobiological correlates of cognitions in fear and anxiety: A cognitive-neurobiological information-processing model. *Cognition & Emotion*, 26, 282–299.

Holmes, A., Nielsen, M. K., Tipper, S., & Green, S. (2009). An electrophysiological investigation into the automaticity of emotional face processing in high versus low trait anxious individuals. *Cognitive, Affective & Behavioral Neuroscience*, 9, 323–334.

Hutchison, W. D., Davis, K. D., Lozano, a M., Tasker, R. R., & Dostrovsky, J. O. (1999). Pain-related neurons in the human cingulate cortex. *Nature neuroscience*, 2, 403–5.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual review of psychology*, 60, 653–70.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, 7, 942–51.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage*, 21, 1167–73.

Ibáñez, A., Hurtado, E., Lobos, A., Escobar, J., Trujillo, N., Baez, S., ... & Decety, J. (2011). Subliminal presentation of other faces (but not own face) primes behavioral and evoked cortical processing of empathy for pain. *Brain Research*, 1398, 72-85.

Ito, T. a, & Urland, G. R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85, 616–26.

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences*, 13, 524–31.

Jabbi, M., Keysers, C. (2008). Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emotion*, 8, 775–80.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage*, 31, 429–39.

Jackson, M. C., Wolf, C., Johnston, S. J., Raymond, J. E., & Linden, D. E. (2008). Neural correlates of enhanced visual short-term memory for angry faces: An FMRI study. *PloS One*, 3, e3536.

Jackson, M. C., Wu, C. Y., Linden, D. E., & Raymond, J. E. (2009). Enhanced visual short-term memory for angry faces. Journal of Experimental Psychology. *Human Perception and Performance*, 35, 363–374.

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience*, 11, 417–28.

Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, 42, 335–46.

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. *The psychology of evaluation: Affective processes in cognition and emotion*, 7-49.

Klumpp, H., Angstadt, M., Nathan, P. J., & Phan, K. L. (2010). Amygdala reactivity to faces at varying intensities of threat in generalized social phobia: An event-related functional MRI study. *Psychiatry Research*, 183, 167–169.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20, 165–182.

Kogan, A., Saslow, L. R., Impett, E. A., Oveis, C., Keltner, D., & Saturn, S. R. (2011). Thin-slicing study of the oxytocin receptor (OXTR) gene and the evaluation and expression of the prosocial disposition. *Proceedings of the National Academy of Sciences*, 108, 19189–92.

Kolassa, I., Kolassa, S., Bergmann, S., Lauche, R., Dilger, S., Miltner, W. H. R., et al. (2009). Interpretive bias in social phobia: An ERP study with morphed emotional schematic faces. *Cognition & Emotion*, 23, 69–95.

Kolassa, I. T., & Miltner, W. H. (2006). Psychophysiological correlates of face processing in social phobia. *Brain Research*, 1118, 130–141.

Kong, J., Gollub, R. L., Polich, G., Kirsch, I., Laviolette, P., Vangel, M., … Kaptchuk, T. J. (2008). A functional magnetic resonance imaging study on the neural mechanisms of hyperalgesic nocebo effect. *The Journal of Neuroscience*, 28, 13354–62.

Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure and Function*, 214, 519–534.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, 11, 539-50.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19, 42–58.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, 54, 2492–502.

Lamm, C., & Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Structure and Function*, 214, 579–91.

Li, W., & Han, S. (2010). Perspective taking modulates event-related potentials to perceived pain. *Neuroscience letters*, 469, 328–32.

Lieberman, M. D. (2013). Social: Why Our Brain Are Wired to Connect.

Lockwood, P. L., Sebastian, C. L., McCrory, E. J., Hyde, Z. H., Gu, X., De Brito, S. A., & Viding, E. (2013). Association of Callous Traits with Reduced Neural Response to Others' Pain in Children with Conduct Problems. *Current Biology*. 23, 901–5.

Lombardo, M. V., Baron-Cohen, S., Belmonte, M. K., & Chakrabarti, B. (2011). Neural endophenotypes of social behavior in autism spectrum conditions. In Decety, J. & Cacioppo, J. T. (Eds.), The Oxford Handbook of Social Neuroscience (pp. 830–847). New York: Oxford University Press.

Luck, S. J. (2008). Visual short-term memory. *Visual Memory*, 43–85.

Luck, S. J., & Hillyard, S. A. (1994a). Electrophysiological correlates of feature analysis during visual search. Psychophysiology, 31, 291–308.

Luck, S. J., & Hillyard, S. A. (1994b). Spatial-filtering during visual-search - evidence from human electrophysiology. *Journal of Experimental Psychology-Human Perception and Performance*, 20, 1000-1014.

Luck, S. J., & Kappenman, E. S. (Eds.). (2011). *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings are accessed but cannot be reported during the attentional blink. *Nature*, 383, 616–618.

Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience*, 22, 496–512.

Maguire, E. A., Mummery, C. J., & Büchel, C. (2000). Patterns of Hippocampal-Cortical Interaction Dissociate Temporal Lobe Memory Subsystems. *Hippocampus*, 10, 475–482.

Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2012). Trust at first sight: Evidence from ERPs. *Social Cognitive and Affective Neuroscience*.

Mathews, A., & MacLeod, C. (1994). Cognitive approaches to emotion and emotional disorders. *Annual Review of Psychology*, 45, 25–50.

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, 1, 167–195.

Mathews, A., & Mackintosh, B. (1998). A cognitive model of selective processing in anxiety. *Cognitive Therapy and Research*, 22, 539–560.

Mathur, V. a, Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *Neuroimage*, 51, 1468–75.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, 36, 455–470.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2012). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*.

Meng, J., Hu, L., Shen, L., Yang, Z., Chen, H., Huang, X., & Jackson, T. (2012). Emotional primes modulate the responses to others' pain: an ERP study. *Experimental Brain Research*, 220, 277–86.

Meng, J., Jackson, T., Chen, H., Hu, L., Yang, Z., Su, Y., & Huang, X. (2013). Pain perception in the self and observation of others: an ERP investigation. *Neuroimage*, 72, 164–73.

Merskey, H., & Bogduk, N. (1994). Classification of chronic pain, IASP Task Force on Taxonomy. *IASP Press, Seattle*.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*. 36, 630–633.

Minio-Paluello, I., Avenanti, A., & Aglioti, S. M. (2006). Left hemisphere dominance in reading the sensory qualities of others' pain?. *Social Neuroscience*, 1, 320–33.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, 27, 237–254.

Morelli, S. A., Rameson, L. T., & Lieberman, M. D. (2012). The neural components of empathy: Predicting daily prosocial behavior. *Social Cognitive and Affective Neuroscience*. 9, 39–47.

Moriguchi, Y., Decety, J., Ohnishi, T., Maeda, M., Mori, T., Nemoto, K., ... & Komaki, G. (2007). Empathy and judging other's pain: an fMRI study of alexithymia. *Cerebral Cortex*, *17,* 2223–34.

Moriguchi, Y., Ohnishi, T., Lane, R. D., Maeda, M., Mori, T., Nemoto, K., Matsuda, H., Komaki, G. (2006). Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. *Neuroimage*, 32, 1472–1482.

Moriya, J., & Sugiura, Y. (2012). High visual working memory capacity in trait social anxiety. *PloS one*, 7, e34244.

Moser, J. S., Huppert, J. D., Duval, E., & Simons, R. F. (2008). Face processing biases in social anxiety: An electrophysiological study. *Biological Psychology*, 78, 93–103.

Nyström, P. (2008). The infant mirror neuron system studied with high density EEG. *Social Neuroscience*, 3, 334–47.

Nyström, P., Ljunghammar, T., Rosander, K., & von Hofsten, C. (2011). Using mu rhythm desynchronization to measure mirror neuron activity in infants. *Developmental Science*, 14, 327–335.

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, 24, 190–198.

Ofan, R. H., Rubin, N., & Amodio, D. M. (2011). Seeing race: N170 responses to race and their relation to automatic racial attitudes and controlled processing. *Journal of cognitive neuroscience*, 23, 3153–61.

Ohnishi, T., Moriguchi, Y., Matsuda, H., Mori, T., Hirakata, M., Imabayashi, E., Hirao, K., Nemoto, K., Kaga, M., Inagaki, M., Yamada, M., & Uno, A. (2004). The neural network for the mirror system and mentalizing in normally developed children: An fMRI study. *Neuroreport*, 15, 1483–1487.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 11087–11092.

Papeo, L., Corradi-Dell’Acqua, C., & Rumiati, R. I. (2011). “She” Is Not Like “I”: The Tie between Language and Action Is in Our Imagination. *Journal of Cognitive Neuroscience*, 23, 3939–3948.

Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nature Reviews Neuroscience*, 2, 417–424.

Pelphrey, K. A., Morris, J. P., & Mccarthy. (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, 16, 1706–1716.

Penolazzi, B., Hauk, O., & Pulvermuller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74, 374-388.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a, Johnson, R., … Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37, 127–52.

Polich, J. (2012). Neuropsychology of P300. *Oxford handbook of event-related potential components*, 159-188.

Preston, S. D., & De Waal, F. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1-20.

Qiu, J. (2005). Pain: Parallel pain pathways. *Nature Reviews Neuroscience*, 6, 827–827.

Raymond, J. E., & O'Brien, J. L. (2009). Selective visual attention and motivation: The consequences of value learning in an attentional blink task. *Psychological Science*, 20, 981–988.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–92.

Rizzolatti, G., Fabbri-destro, M., & Cattaneo, L. (2009). Mirror neurons and their clinical relevance. *Nature clinical practice neurology*, 5, 24–34.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11, 264-274.

Rossignol, M., Philippot, P., Bissot, C., Rigoulot, S., & Campanella, S. (2012). Electrophysiological correlates of enhanced perceptual processes and attentional capture by emotional faces in social anxiety. *Brain Research*, 1460, 50–62.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature neuroscience*, 4, 546–50.

Ruby, P., & Decety, J. (2003). What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *European Journal of Neuroscience*, 17, 2475–2480.

Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of Cognitive Neuroscience*, 16, 988–99.

Rudoy, J. D., & Paller, K. A. (2009). Who can you trust? behavioral and neural differences between perceptual and memory-based influences. *Frontiers in Human Neuroscience*, 3, 16.

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition*, 125, 207–218.

Saarela, M. V, Hlushchuk, Y., Williams, A. C. D. C., Schürmann, M., Kalso, E., & Hari, R. (2007). The compassionate brain: humans detect intensity of pain from another's face. *Cerebral Cortex*, 17, 230–7.

Said, C. P., Baron, S. G., & Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience*, 21, 519–528.

Santos, I. M., & Young, A. W. (2008a). Effects of inversion and negation on social inferences from faces. *Perception*, 37, 1061–1078.

Santos, I. M., & Young, A. W. (2008b). Exploring the perception of social characteristics in faces using the isolation effect. *Visual Cognition*, 12, 213–247.

Saxe, R, & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind". *Neuroimage*, 19, 1835–1842.

Saxe, Rebecca. (2005). Against simulation: the argument from error. *Trends in Cognitive Sciences*, 9, 174–9.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, 43, 1391–1399.

Sessa, P, Luria, R., Verleger, R., & Dell'Acqua, R. (2007). P3 latency shifts in the attentional blink: further evidence for second target processing postponement. *Brain Research*, 1137, 131–9.

Sessa, P., Luria, R., Gotler, A., Jolicoeur, P., & Dell'acqua, R. (2010). Interhemispheric ERP asymmetries over inferior parietal cortex reveal differential visual working memory maintenance for fearful versus neutral facial identities. *Psychophysiology*, 48, 187–197.

Sessa, P., Tomelleri, S., Luria, R., Castelli, L., Reynolds, M., & Dell'Acqua, R. (2012). Look out for strangers! Sustained neural activity during visual working memory maintenance of other-race faces is modulated by implicit racial prejudice. *Social Cognitive and Affective Neuroscience*, 7, 314–321.

Shamay-Tsoory, S.G., Tomer, R., Goldsher, D., Berger, B.D., Aharon-Peretz, J. (2004). Impairment in cognitive and affective empathy in patients with brain lesions: anatomical and cognitive correlates. *Journal of Clinical and Experimental Neuropsychology*, 26, 1113–27.

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132, 617–27.

Sheng, F., & Han, S. (2012). Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. *Neuroimage*, 61, 786–97.

Sheng, F., Liu, Y., Zhou, B., Zhou, W., & Han, S. (2013). Oxytocin modulates the racial bias in neural responses to others ' suffering. *Biological Psychology*, 92, 380–386.

Schulte-Ruther, M., Markowitsch, H.J., Fink, G.R., Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, 19, 1354–72.

Sica, C., Musoni, I., Chiri, L.R., Bisi, B., Lolli, V., Sighinolfi, C. (2007). Social Phobia Scale (SPS) e Social Interaction Anxiety Scale (SIAS): traduzione ed adattamento italiano. *Bollettino di Psicologia Applicata*, 252, 59.

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2008). Levels of emotional awareness and autism: an fMRI study. *Social neuroscience*, 3, 97–112.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, 30, 855-863.

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in cognitive sciences*, 13, 334–340.

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, 653–62.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences,* 1156, 81–96.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303, 1157–62.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439, 466–9.

Spielberger, C.D. (1983). Manual for the State-Trait Anxiety Inventory: STAI (Form Y). Palo Alto, CA: Consulting Psychologists Press.

Spiers, H. J., & Maguire, E. A. (2006). Spontaneous mentalizing during an interactive real world task: an fMRI study. *Neuropsychologia*. 44, 1674–1682.

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological science*, 21, 1593–8.

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *Neuroimage*, 59, 3050–9.

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. Journal of cognitive neuroscience, 23, 63–74.

Stein, Edith (1964). On the Problem of Empathy. The Hague: Martinus Nijhoff. (Orig. pub. 1917.)

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, 21, 349–354.

Stout, D. M., Shackman, A. J., & Larson, C. L. (2013). Failure to filter: Anxious individuals show inefficient gating of threat from working memory. *Frontiers in Human Neuroscience*, 7, 1–9.

Sullivan, M. J. L., Martel, M. O., Tripp, D. a, Savard, a, & Crombez, G. (2006). Catastrophic thinking and heightened perception of pain in others. *Pain*, 123, 37–44.

Telzer, E. H., Masten, C. L., Berkman, E. T., Lieberman, M. D., & Fuligni, A. J. (2011). Neural regions associated with self control and mentalizing are recruited during prosocial behaviors towards the family. *Neuroimage*, 58, 242–9.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, 1124, 208–224.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience*, 3, 119–127.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*, 455–460.

Todorov, A., Said, C. P., Oosterhof, N. N., & Engell, A. D. (2011). Task-invariant brain responses to the social value of faces. *Journal of Cognitive Neuroscience*, 23, 2766–2781.

Umiltà, M. a, Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing. a neurophysiological study. *Neuron*, 31, 155–65.

van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241-255.

Van Overwalle, F. (2009). Social Cognition and the Brain : A Meta-Analysis. *Human Brain Mapping*, 30, 829–858.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*, 564–584.

Verosky, S. C., & Todorov, A. (2010). Differential neural responses to faces physically similar to the self as a function of their valence. *Neuroimage*, 49, 1690–1698.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428, 748–751.

Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *The Journal of neuroscience*, 32, 7646–50.

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: social-cognitive goals affect amygdala and stereotype activation. *Psychological science*, 16, 56–63.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. Neuron, 40(3), 655–64.

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological science*, 17, 592–8.

Wilkinson, D. T., & Halligan, P. W. (2004). The relevance of behavioural measures for functional imaging studies of cognition. *Nature Reviews Neuroscience*, 5, 67–73.

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, 277–283.

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *The Journal of Neuroscience*, 29, 8525-8529.

Yang, D., Qi, S., Ding, C., & Song, Y. (2011). An ERP study on the time course of facial trustworthiness appraisal. *Neuroscience Letters*, 496, 147–151.

Yoon, K. L., & Zinbarg, R. E. (2007). Threat is in the eye of the beholder: Social anxiety and the interpretation of ambiguous facial expressions. *Behaviour Research and Therapy*, 45, 839–847.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience*, 15, 675–680.

Zaki, J., & Ochsner, K. (2009). The need for a cognitive neuroscience of naturalistic social cognition. *Annals of the New York Academy of Sciences*, 1167, 16–30.

Zebrowitz, L. A., & Montepare, J. M. (2005). Psychology. appearance DOES matter. *Science*, 308, 1565–1566.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, 2, 1497.

Gli anni del dottorato sono stati fra i più intensi della mia vita. I momenti di sconforto si sono alternati a momenti di pura gioia e in mezzo c'è stata l'intera gamma delle emozioni umane.

Alcuni colleghi con cui ho condiviso questo percorso sono diventati cari amici e sono certa che siano legami difficili da rompere, proprio per la grande varietà di emozioni che hanno caratterizzato questi anni. Fare ricerca ci spinge inevitabilmente ai margini di noi stessi rendendoci vulnerabili, perché ci mettiamo tutti noi stessi e spesso c'è bisogno di rassicurazioni che spesso, indirettamente, queste persone mi hanno dato, con loro mi sono sempre sentita accolta.

Paola ha supervisionato il mio percorso instaurando con me un rapporto di grande umanità, credendo in me anche quando io non riuscivo a farlo, dando tutto lo spazio possibile alle mie idee, spesso fantascientifiche. Amica e guida. E così anche Roy, mentore di sempre.

Vorrei infine esprimere tutta la mia gratitudine per la mia famiglia e in particolare mamma, papà e Sara, naturalmente, per il loro entusiasmo. Per essersi sempre entusiasmati, e per le grandi risate, a casa nostra si è sempre riso tantissimo, grazie per questo anche ai miei cari zietti, nonnina e Simona. Questo ha davvero riempito tutti gli aspetti della mia vita e anche nella ricerca questo mi ha ispirata sempre.

Infine Matteo. Perché tutto si è illuminato allora. Perché siamo la nostra opera d'arte.

"stretta la foja, largo er viale

pijateve la favola come ve pare"