The speed of morality: a high-density electrical neuroimaging study

Jean Decety¹,² and Stephanie Cacioppo¹,³

¹. Department of Psychology - The University of Chicago, 60637, Chicago, IL
². Department of Psychiatry and Behavioral Neuroscience
³. Department of Psychology, University of Geneva, 1205, Geneva, Switzerland

Abstract

Neuroscience research indicates that moral reasoning is underpinned by distinct neural networks including the posterior superior temporal sulcus (pSTS), amygdala, and ventromedial prefrontal cortex, which support communication between computational systems underlying the affective states, cognitions, and motivational processes. To characterize real time neural processing underpinning moral computations, high-density event-related potentials were measured in participants while they viewed short morally-laden visual scenarios depicting intentional and accidental harmful actions. Current source density maxima in the right pSTS as fast as 62 ms post-stimulus first distinguished intentional vs. accidental actions. Then responses in the amygdala (122 ms) and ventromedial prefrontal cortex (182 ms) were evoked by the perception of harmful actions, indicative of fast information processing associated with early stages of moral cognition. Our data strongly supports the notion that intentionality is the first input to moral computations. They also indicate that emotion acts as a gain antecedent to moral judgment by alerting the individual of the moral salience of a situation, and provide evidence for the pervasive role of affect in moral sensitivity and reasoning.

Key words: moral cognition, emotion, intention understanding, high-density ERPs, posterior superior temporal sulcus (pSTS), amygdala, ventromedial prefrontal cortex.
Introduction

Moral decision-making is a fundamental aspect of social cognition, and is seen as a product of our biological evolutionary and cultural history. In the past decade, research in multiple academic domains including evolutionary biology, developmental science, economics and cognitive neuroscience has endeavored to more clearly define and investigate this construct. Results from these disciplines suggest that mature moral abilities emerge from a sophisticated integration of cognitive, emotional and motivational mechanisms (Decety & Howard, 2012). For instance, developmental research indicates that very early in life (around 3-month of age), pre-verbal babies express what appear to be nascent moral evaluations (Hamlin et al., 2011), and that toddlers differentially interpret intentional from accidental or non-purposeful actions (Woodward & Sommerville, 2000). In judging whether another individual’s actions are good or bad, harmful or harmless, our capacity to recognize a behavior as intentional is a central component of human social cognition (Cushman, Young & Greene, 2010; Malle & Guglielmo, 2012).

Early indications that moral judgment depends on specific neural substrates come from neurological observations and case studies, such as the emblematic case of Phineas Gage, a railroad worker who’s decision making in real life was profoundly changed and impaired after massive, yet localized damage of his ventromedial prefrontal cortex (Prehn & Heekeren, 2009). Subsequent studies from patients with such lesions support the critical role of this region in moral decision-making and moral behavior (e.g., Gleichgerrcht et al., 2011; Koenigs et al., 2007; Moretto et al., 2010). Interestingly, impairments of the ventromedial prefrontal cortex (vmPFC) are not associated deficits in mental states understanding.

In the past decade, functional MRI (fMRI) has begun to identify a network of brain regions involved in moral cognition (Moll & Schulkin, 2009; Young and Dungan, 2012 for recent reviews). These studies indicate that a restricted number of regions are involved in moral reasoning, including the posterior superior temporal sulcus
(pSTS, also reported in the literature as the TPJ), amygdala, and vmPFC (Borg Schaich et al., 2011; Decety & Porges, 2011; Greene et al., 2001; Heekeren et al., 2003, Moll et al., 2002, 2007). Moral reasoning seems to be underpinned by specific neural circuitry, but, in fact, these circuits are not unique to morality; rather, they involve regions and systems underlying specific affective states, cognitive and motivational processes.

One serious limitation of a majority of previous studies is that the results rely only on subtraction logic in their designs, which is based on the a priori assumption that one computational process can be added to a pre-existing set of processes without affecting them, assuming that there are no interactions among the different components of a given task. Furthermore, characterizing brain activity in terms of functionally segregated regions does not reveal anything about how different brain regions communicate with each other. Connectivity analyses and high-density event-related potentials (ERPs) can identify patterns of communication between regions that contrast analyses may not detect. Such methods are necessary to advance of knowledge on the neuroscience of morality. This is also important at a theoretical level. The fact that fMRI studies found activation in emotion-related areas such as the amygdala, and vmPFC during morally salient stimuli provides only correlational data, showing that emotions are associated with moral cognition, but is insufficient to determine whether affective processing is taking place during moral evaluations or antecedent to them (Huebner, Dwyer & Hauser, 2008). In addition, there is no information on the timing and order of component processes involved in moral judgment.

In the current study, we took advantage of high-density ERPs and analysis of source localizations to examine the spatio-temporal dynamics of the neural processing evoked by the perception of visual morally-laden scenarios in healthy individuals. We used a well-established paradigm that was validated with fMRI and eye-tracking measures in a large number of participants aged between 4 to 37 years (Decety, Michalska & Kinzler, 2012). The results of that study showed that the perception of
intentional harm versus accidental harm was associated with increase signal in the amygdala, periaqueductal gray, insula, vmPFC, and right pSTS/TPJ. High-density event-related potentials can identify patterns of communication between regions that contrast analyses may not detect and such methods are necessary to advance of knowledge of the neuroscience of morality.

**Methods**

**Participants:** A total of ten healthy participants were involved in the study. All provided written informed consent to participate in the experiment, which was approved by the local Committee for Protection of Human Subjects. Three volunteers had to be excluded due to artifacts in their EEG data. Thus, brain activity from seven (4 females; 3 males) volunteers was fully analyzed. All participants were right-handed, older than 18 (mean age: 21.86; SD: 3.13), native English speakers with normal or corrected-to-normal vision, and were not taking antidepressant medication. None of the participants had prior or current neurological or psychiatric disorders (e.g., traumatic brain injury with loss of consciousness, epilepsy, neurological impairment or degenerative neurological illness), as ascertained by a detailed anamnesis.

**Behavioral procedure:** Participants completed a modified version of a standard Intention Inference Task (IIT) developed by Decety, Michalska and Kinzler (2012) in studies on empathy and morality, according to a 2 x 2 factorial design with intention type (intentional vs. accidental) and target type (object vs. person) as within-subjects factors. While participants’ electrical brain activity was recorded, participants were required to watch the stimuli, to gaze at the center of the screen, and to judge whether the action was performed intentionally or accidentally. During this part of the experiment, no reaction times were collected to avoid any motor artifacts. After the completion of the EEG recordings, one additional behavioral block was run, during which accuracy and reaction times were recorded.
Behavioral paradigm: During the IIT, participants watched a series of three frame-video clips that were presented centrally on a monitor screen, as in a previous functional MRI study (Decety et al., 2012). The first frame (T1) from the video-clip was 500 ms long and displayed an establishing scene, the second frame (T2) was a 700 ms frame displaying either an intentional harm or an accidental harm, followed by a third 1000 ms frame (T3) confirming the intentional or accidental harm. The three-frame video-clip technique provided a tight control of the kinematics, task duration, and timing. Each trial began with a fixation cross that was presented for 150 ms. A 1500 to 2000 ms maximum inter-trial interval was randomly inserted.

Instructions: Participants were instructed to gaze at the center of the screen and to judge whether the action was performed intentionally or accidentally. In addition, participants were asked to refrain from blinking or moving their eyes except during the interval between trials.

Electrophysiological recordings: Continuous EEG was recorded from 128 AgCl carbon-fiber coated electrodes using an Electric Geodesic Sensor Net® (GSN300; Electrical Geodesic Inc., Oregon; http://www.egi.com/), where EEG electrodes are arrayed in a dense and regular distribution across the head surface with an inter-sensor distance of approximately 3 cm. The EEG was digitized at 500 Hz (corresponding to a sample bin of 2 ms), band-width at 0.01–200 Hz, with the vertex electrode (Cz) serving as an on-line recording reference; impedances were kept below 50 kΩ. Participants were seated in a comfortable chair 150 cm away from a PC computer screen in which stimuli were presented centrally.

Electrophysiological data processing: Electrophysiological data were imported and analyzed in Cartool (version 3.32; Denis Brunet; http://brainmapping.unige.ch/Cartool.htm). First, epochs of analysis were visually inspected for oculomotor (saccades, and blinks), muscles, and other artifacts in addition to an automated threshold rejection criterion of 100 microvolts.
After off-line artifact rejections, VEPs were computed from 0 to 400 ms after the onset of T2 (i.e., the second frame displaying either an intentional harmful action or an accidental harmful action). Visual evoked potential (VEP) data were then band-pass filtered between 1 and 30 Hz without baseline correction. VEP data were next recalculated off-line against the average reference and normalized to their mean global field power (i.e., GFP) before group-averaging. The GFP is computed as the spatial standard deviation of the scalp electric field, yields larger values for stronger electric fields, and is calculated as the square root of the mean of the squared value recorded at each electrode (vs. the average reference). Channels with corrupted signals and channels showing substantial artifacts and noise throughout the recordings were interpolated to a standard 111-channel electrode array using a three-dimensional spline procedure.

**Second level electrical data analysis:** Topographical analyses: VEP data were analyzed with space-oriented brain electric field analysis using Cartool. Because this electrophysiological method has been extensively detailed previously (see Brunet, Murray & Michel, 2011), here we provide only the essentials details. This space-oriented brain electric field approach is based on the empirical observation that the brain electric field configuration changes step-wise overtime. Epochs of quasi-stable field configurations ('microstates’) are concatenated by abrupt transitions in the brain electric field configurations. Microstates thus are assumed to implement specific brain functions. To identify start and end of each optimal microstate, a standard cluster analysis previously described was employed using the grand-mean ERPs of each condition. This cluster analysis uses a hierarchical agglomerative cluster-algorithm to identify the predominant topographies (i.e., maps) and their sequence within a data set (these methods are implemented in Cartool). The optimal number of maps (i.e., the minimal number of maps that accounts for the greatest variance of the data set) is determined based on a modified Krzanowski–Lai criterion. Importantly, this pattern analysis is reference-free and insensitive to amplitude modulation of the same scalp potential field across conditions, since normalized maps are compared. We also applied the constraint
that a given scalp topography must be observed for at least five consecutive data points (i.e., 10 ms at a 500-Hz digitization rate) in the group-averaged data. This criterion is effectively similar to that frequently applied in the analysis of VEP waveform modulations. This pattern analysis was performed across time and experimental conditions in order to determine whether and when different conditions recruited different configurations of intracranial generators. This conservative approach allowed us to analyze the VEP’s responses for intentional vs. accidental harmful actions. Then, the pattern of maps observed in the group-averaged data was statistically tested by comparing each of these maps with the moment-by-moment scalp topography of individual subjects’ VEP from each condition. To do so, each time point of each VEP from each subject was labeled according to the map with which it best correlated spatially. In other words, the optimal number of maps was fitted into the original data for each individual subject, using a competitive fitting procedure. This “fitting” procedure determines whether a given experimental condition is more often described by one map versus another. From this “fitting” procedure, a large amount of information can be extracted and analyzed for a given condition across subjects, such as the total amount of time a given stable configuration. This latter value represents the frequency with which a given microstate was observed within a given time period for each experimental condition. This is the information used here. Then, the extracted values of interest were subjected to a repeated-measure ANOVA. Results were accepted as significant at p < 0.05. The hierarchical cluster analysis of the VEP topographies in this “Action type” condition revealed a total of six time periods of stability that significantly differed both in timing and location for intentional harm and accidental harm in the 400 ms post-stimulus period that explained 90.14 % of variance in the collective data set (Figure 1). The windows of occurrence for these stable time periods corresponded to the following time intervals: Microstate 1 (M1): 0–20 ms; Microstate 2 (M2): 22–60 ms; Microstate 3 (M3): 62–120 ms for intentional actions; and from 62 to 140ms for accidental actions; Microstate 4 (M4): 122-180 ms for intentional actions and 142 to 180 ms for accidental actions; Microstate 5 (M5): 182-274 ms for intentional actions and 182 to 304 ms for accidental actions;
Distributed intracranial source estimations: The intracranial generators for each condition were estimated using a distributed linear inverse solution based on a low-resolution brain electromagnetic tomography (LORETA) model of the unknown current density in the brain. Here, LORETA was used with a lead field (solution space) that was calculated on a realistic head model that included 3005 solution points, selected from a 6 x 6 x 6 mm grid equally distributed within the gray matter. Source estimations were rendered on the MNI/McGill average standard brain as supplied by Cartool. Accuracy of anatomical labeling was ascertained with a visual inspection of the Duvernoy (1991) brain atlas. The maximum current density for each microstates were the following: M1: 3, 44, -7 (xyz Talairach coordinates); M2: 52, -60, 23; M3: 52, -60, 21; M4: 32, 0, -30; M5: 3, 38, -7; M6: 3, 38, -7.

Results and Discussion

To determine the timing and order of component processes implicated in moral cognition and whether affective processing occurs during moral evaluations or antecedent to them, we used high-density ERPs to examine the spatio-temporal dynamics of the neural processing evoked by the perception of visual morally-laden scenarios. The perception of intentional harm was associated with better (90% vs. 71%) and faster reaction times for intentional harm compared to accidental harm ($p < 0.05$), as well as the specific involvement of the right pSTS, amygdala and vmPFC. These regions were found activated in an fMRI study using the same stimuli and same contrast (Decety et al., 2012). Interestingly, intentional harmful actions were significantly distinguished from the accidental harmful actions in three main time periods (i.e., from 62 to 140 ms, and from 122 to 180 ms post-stimulus, see Figure 1). More precisely, a specific scalp potential field with a current source density maximum in the right pSTS (x52, y-60, z21; Talairach coordinates) characterized
accidental harm during the first time period, whereas a different scalp topography
with a current source density in the right amygdala (32, 0, -30, xyz Talairach
coordinates from 122-180ms) and vmPFC (3, 38, -7, xyz Talairach coordinates from
182-304 ms) characterized the perception of intentional harm (Table 1).

As expected from previous work on the perception of intentions (Grafton, 2009;
Ortigue et al., 2009), the first significant brain microstate emerged around 60 ms
over the posterior STS. Interestingly, this microstate remained for a longer period
for scenarios depicting accidental actions (from 62 to 140 ms; Figure 2), a finding
consistent with the involvement of this region in the visual analysis of other people's
actions and intentions. Activity in the pSTS is greater for incongruent than for
congruent actions, demonstrating a need for different levels of processing for
observed goal-directed and non-goal-directed observed actions (Pelphrey & Carter,
2008). The STS region, which receives input from both the ventral and dorsal visual
pathways and which projects to the amygdala and vmPFC, plays a critical role in the
visual analysis of perceptual information about bodily motion, action prediction, and
evaluation of the intentions behind other people's behaviors (Allison, Puce &
MacCarthy, 2000). Previous studies have shown the presence of recruitment in STS
and intraparietal sulcus early in visual processing in line with recent
electrophysiological evidence from both animal and human studies arguing for a
bidirectional mechanism where these associative areas can be recruited in very
early stages of information processing (~60 ms) (Ortigue et al., 2009). The fast
involvement of the right pSTS in differentiating intentional from accidental harm
demonstrates that automatic perception of intentionality is a critical input to the
perception of moral valence of an action. This component, both in timing and
localization, is different from emotional negativity bias and stimulus evaluation that
are associated with later responses (Huang & Luo, 2006). Previous research using
ERPs and magnetoencephalography has reported a frontal N110-140 and a late
centro-parietal P300 when individuals viewed faces and body parts being injured
(Chen, Yang, & Cheng, 2012; Cheng et al., 2008; Decety, Yang & Cheng, 2010; Fan &
Han, 2008). The early frontal negativity component, 110-140 ms after visual
stimulation, whose source density originates in the dorsal anterior cingulate cortex, is interpreted as an automatic attentional response, whereas the P300 reflects stimulus evaluation and classification.

The early engagement of the right amygdala and ventromedial prefrontal cortex evoked by the perception of intentional harmful actions suggest that affective processes precede cognitive evaluative processes. The amygdala through reciprocal connection with the pSTS underlies rapid and prioritized processing of emotion signals (Sander et al., 2003). The vmPFC projects to the basal forebrain (the major cholinergic output) and brainstem regions, which contains all afferent and efferent systems necessary for survival, including basic affective responses (Ongür & Price, 2000), and neurons within the vmPFC encode the emotional value of stimuli (Elliott et al., 2010). Further, the vmPFC, reciprocally connected with the amygdala and hypothalamus, is involved in the autonomic component of emotion, and seems essential in the evaluation of harmful intent. Dysfunction within these regions or their functional connectivity has reliably been associated with impaired moral decision making in psychopathy (Anderson & Kiehl, 2011; Blair, 2007).

While previous neuroimaging research indicates that emotional processing is an integral aspect of moral cognition, the exact point at which this effect occurs has not yet been examined. Here, taking advantage of the high-density ERPs, coupled with brain source analysis, and building on a paradigm validated with fMRI, we demonstrate for the first time how intention understanding in the right pSTS (62 ms after stimulus onset), and then affective processing occurs in very early stages of moral cognition processing (i.e., 122-180 ms after stimulus onset over the amygdala and 182 ms in the vmPFC). These results support the view that intentionality judgments both precede and guide moral cognition (Malle & Guglielmo, 2011).

The timing of early engagement of the amygdala and vmPFC during the perception of intentional harm is consistent with the notion that emotion acts as a gain antecedent to moral judgment by alerting the individual of the moral salience of a
situation and provides evidence for the pervasive role of affect in moral cognition (Blair & Fowler, 2008). Finally, our results are also important for neurodevelopmental research with typical and atypical populations, such as children with callous-unemotional traits, which indicates that affective arousal is a necessary (although not sufficient) precursor to mature empathic understanding and moral decision-making (Blair, 1995; Cheng et al., 2012; Decety, Michalska & Kinzler, 2011).

A limitation of the present study is that it included only a small number of participants. Replication with a greater number of subjects will increase our confidence in the generalizability of these findings.

Acknowledgments: We thank Tee Zhou for technical help. JD was supported by a grant from NSF (BCS-0718480) and SC by a grant from the Swiss National Science Foundation (FNS_PP00_1_128599/1).

References


Table 1: Local maxima of current source density obtained from LORETA brain source estimations of EEG data for intentional harmful actions and accidental harmful actions.

Figure 1: Visual event-related potentials (VEPs) and brain microstates. Group-averaged VEPs elicited by the presentation of visual stimuli depicting intentional harm (Left), and accidental harm (Right). The topographic cluster analysis identified three main distinct brain microstates (colored bars) in the 400 ms following stimulus (T2) presentation. These three brain microstates (blue, green and yellow bars, respectively) had different duration depending upon the stimulus type (see Table 1 for further details).

Figure 2: Schematic representation of the timing and anatomical localization of brain microstates in response to accidental harm (top-left panel) in the right pSTS/TPJ (62-140 ms) and to intentional harm (bottom-left panel) in the right amygdala (122-180 ms) and ventromedial prefrontal cortex (182-304 ms). Stimulus exemplars of the two classes of stimuli (intentional and accidental harmful actions) are shown on the right side of the panel. Three transverse brain sections showing the estimated localization of the intracranial brain generators of the three main microstates.
**Table 1**: Local maxima of current source density obtained from LORETA brain source estimations of EEG data for intentional harmful actions and accidental harmful actions.

<table>
<thead>
<tr>
<th>Microstate time periods for <em>Intentional</em> Harmful Actions</th>
<th>Microstate time periods for <em>Accidental</em> Harmful Actions</th>
<th>Brain Microstate Label</th>
<th>Talairach coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-20 ms</td>
<td>0-20 ms</td>
<td>M1</td>
<td>3 44 -7</td>
</tr>
<tr>
<td>22-60 ms</td>
<td>22-60 ms</td>
<td>M2</td>
<td>52 -60 6</td>
</tr>
<tr>
<td>62-120 ms</td>
<td>62-140 ms</td>
<td>M3</td>
<td>52 -60 11</td>
</tr>
<tr>
<td>122-180 ms</td>
<td>142-180 ms</td>
<td>M4</td>
<td>32 0 -30</td>
</tr>
<tr>
<td>182-274 ms</td>
<td>182-304 ms</td>
<td>M5</td>
<td>3 38 -7</td>
</tr>
<tr>
<td>276-400 ms</td>
<td>306-400 ms</td>
<td>M6</td>
<td>3 38 -7</td>
</tr>
</tbody>
</table>