Early categorization of social affordances during the visual encoding of bodily stimuli.

Authors: Moreau Q. [1,2]*, Parrotta E.[3], Pesci, U.G.[1,2], Era V.[1,2], Candidi M[1,2]*.

1. Department of Psychology, Sapienza University, Rome, Italy
2. IRCCS Fondazione Santa Lucia, Rome, Italy
3. School of Psychology, Action Prediction Lab, University of Aberdeen, Aberdeen, Scotland

*Corresponding authors: QM at quentin.moreau@ppsp.team, MC at matteo.candidi@uniroma1.it

Funding: MC was supported by the Italian Ministry of Health (Ricerca Finalizzata, Giovani Ricercatori 2016, n. GR-2016-02361008) and Sapienza University (Progetti di Ricerca Grandi 2020, RG12017B8343252). VE was supported by the Fondazione Umberto Veronesi.

Keywords: Social affordance; Action perception; Hands; EEG; MVPA

Abstract

Interpersonal interactions rely on various communication channels, both verbal and non-verbal, through which information regarding one's intentions and emotions is transmitted. Here, we investigated the neural correlates underlying the visual processing of hand postures conveying social affordances (i.e., hand-shaking), compared to control stimuli such as hands performing non-social actions (i.e., grasping) or showing no movement at all. Combining univariate and multivariate analysis on electroencephalography (EEG) data, our results show that occipito-temporal electrodes show an early (before 200 ms after image presentation) differential processing of stimuli conveying social information compared to non-social ones. First, the amplitude of two Event-Related Potentials related to the perception of body parts (i.e., the N190 and the Early Posterior Negativity - EPN) is modulated differently during the perception of social and non-social content carried by hands. Moreover, our multivariate classification analysis (MultiVariate Pattern Analysis - MVPA) expanded the univariate results by revealing two crucial time windows for distinguishing implied motion with and without social affordances over occipito-parietal sites. In conclusion, we provide new evidence suggesting that the encoding of socially relevant hand gestures is categorized in the early stages of visual processing and future studies will need to clarify whether and how fronto-parietal networks contribute to structuring this functional organization in the visual system.

Introduction

Humans spend much of their lives observing how their conspecifics act, predicting and understanding their actions and intentions, and learning from them. These reciprocal processes are crucial to efficient social interactions between individuals. Indeed, during interpersonal interactions, non-verbal information conveyed by facial expressions, body postures, and hand movements represents a key channel through which others express their intentions and their willingness to make connections (Argyle, 1972).

Based on evidence from neurophysiological (Thierry et al., 2006; Meeren et al., 2013), neuroimaging (fMRI, Downing et al., 2001; Gschwind et al., 2012), brain lesion (Sliwinska et al., 2020), and brain stimulation studies (Urgesi et al., 2007; Sadeh et al., 2011), a specific visual pathway for the processing of social stimuli such as faces, bodies, and their movements, has recently been proposed (Pitcher and Ungerleider, 2021). Located on the lateral surface of the occipital and parietal lobes, this third-visual "social" pathway is an addition to the two-streams model (i.e., the "what" ventral pathway and the "where" dorsal pathway, Milner and Goodale, 2006; Milner and Goodale, 2008; Goodale, 2014).

While most seminal neurocognitive models have focused on face perception and its contribution to sociality (see Haxby et al., 2000; Haxby et al., 2002; Hugenberg and Wilson, 2013 for reviews), body postures and hand gestures also constitute a highly relevant source of social information for efficient interpersonal interactions (Ramsey, 2018). The processing of body images has been associated with the activity of a bilateral occipito-temporal focal region, the Extrastriate Body Area (EBA; Downing et al., 2001), whose causal role in visual body perception has been confirmed by brain stimulation and lesion studies (Urgesi et al., 2004; Urgesi et al., 2007; Moro et al., 2008; Moro et al., 2012; Gandolfo and Downing, 2019). Interestingly, imaging studies showed a topographical functional organization of high-order temporo-occipital visual areas (Orlov et al., 2010), with specific brain activity associated with the processing of hand images (compared to other body parts) in the lateral occipital-temporal cortex and the extrastriate body area (i.e., LOTC and EBA; Bracci et al., 2010; Bracci et al., 2012). Patterns of activity of the LOTC are thought to form a representational space encompassing low- (e.g., body-related shapes) and high-level (action type, social meaning) aspects of social vision (Lingnau and Downing, 2015; Wurm *et al.*, 2017; Tucciarelli *et al.*, 2019; Tarhan and Konkle, 2020). Non-invasive brain stimulation and lesion studies showed that the discrimination and recognition of implied actions further depend on the activity of superior temporal (Candidi *et al.*, 2011; Urgesi *et al.*, 2014; Candidi *et al.*, 2015), parietal (Cattaneo *et al.*, 2010; Avenanti *et al.*, 2013), and premotor (Urgesi *et al.*, 2007; Candidi *et al.*, 2008; Cattaneo *et al.*, 2010; Avenanti *et al.*, 2013; Urgesi *et al.*, 2014) regions.

The role of occipito-temporal areas in action perception is currently debated. Especially the pending interrogation remains to what extent the LOTC alone can handle the encoding and the discrimination of actions, and how this activity is later supported/integrated within the fronto-parietal network (Lingnau & Downing, 2015; Tucciarelli et al., 2019; Tarhan et al., 2020). Indeed, to the best of our knowledge, only one MEG study (Tucciarelli et al., 2015) investigated the spatio-temporal dynamics of action discrimination and indicated that the activity of lateral occipito-temporal regions (i.e., consistent with the EBA) has the earliest access (compared to precentral regions) to the discrimination of different observed actions (e.g. grasping and pointing) independently from their low-level features. Yet, no study has tested the dynamics of social action representation, crucial to the understanding of how social affordance are encoded by the visual cortex.

The response of occipito-temporal areas to the presentation of body (hand) images has been characterized in the temporal domain by Magneto- Electro-, and intracortical-physiological recording studies (i.e., M/EEG/ECoG) showing that the visual perception of hands modulates the amplitude of a negative component (i.e., the N190) over occipito-temporal regions (Thierry et al., 2006; Ishizu et al., 2010). Interestingly, the perception of hand images elicits a higher amplitude of the N190 compared to both whole bodies and control images (Espírito Santo et al., 2017; Moreau et al., 2018; Moreau et al., 2019). Besides these early and posterior brain responses to visual bodily stimuli, EEG studies have also highlighted the sensitivity of later and more cognitive brain responses such as the Early Posterior Negativity (EPN), peaking around 300 ms after stimulus presentation, to salient stimuli. So far, EPN has been studied during the visual processing of socially relevant stimuli such as emotional faces (Sato et al., 2001) and body postures (Borhani et al., 2015) and its modulation is thought to reflect attentional selection to relevant stimuli that will generate further processing (Schupp, Junghöfer, et al., 2004).

In the present study, we aimed to describe the neural dynamics associated with perceiving hand stimuli that convey social affordances. We recorded EEG activity during the passive viewing of hand pictures either being still, performing a non-social action (i.e., grasping), or conveying a social affordance (i.e., hand-shaking; Figure 1A). Our main hypothesis was that the visual perception of hands conveying social affordances would generate different EEG signatures (i.e., N190 and EPN) relative to hands not providing any social affordance. To do so, classical univariate (i.e., event-related potentials - N190 and EPN) and single-trial multivariate (Multivariate Pattern Analysis - MVPA) analyses were combined to explore the processing of socially relevant hand postures. This approach has recently gained momentum to get new insights on the neural patterns of activity underlying perceptual and cognitive processes across multiple scales (Grootswagers et al., 2017)

Material and Methods

Participants

23 healthy individuals (10 males and 13 females, mean age 25.3 years) took part in the experiment. The sample size was established using a power analysis, performed with the software More Power (Campbell and Thompson, 2012). We inserted as expected effect size the partial eta squared value obtained by Moreau et al. 2019 ($\eta_p^2 = 0.29$). The output indicates that a three-level factor in a within-subject design, with a power of 0.95 and a partial eta squared of 0.29, requires a sample size of 22 participants.

All participants were right-handed with normal or corrected-to-normal vision acuity in both eyes. Information about the purpose of the study was provided only after the experimental tests were completed. Participants gave their written informed consent and were reimbursed at a rate of 7€/hour. The experimental procedures were approved by the Ethical Committee of the Fondazione Santa Lucia (Rome, Italy) and were in accordance with the ethical standards of the 2007 Declaration of Helsinki.

Stimuli

Stimuli comprised a set of 12 grayscale pictures taken with a digital camera in the same space and light intensity-controlled conditions for each class of stimuli. The stimuli set was made of three types of hand pictures: 1) Still hands (no action), 2) Grasping hands conveying an implied action without social affordance, and 3) Shaking hands conveying an implied action with social affordances (Figure 1A). Each class comprised a set of four images of volunteers' right hands (2 males, 2 females). The images were modified by means of the Adobe Photoshop software to obtain a set matched for dimension, orientation, background, contrast, and luminosity; all images were scaled to 400x400 pixels at a resolution of 72 dpi on 'Netscape grey' (128;128;128).

Stimuli have been validated by two online surveys beforehand (Survey Monkey) to control for both clarity (distinguishing the specific movement) and socially informative features (distinguishing the presence/absence of social cues), as well as rates of implied motion conveyed by the stimuli (Grasping and Shaking hands). In the first survey, 30 individuals were required to evaluate the original set of 22 images, by assessing the degree of implied motion conveyed by each image (Likert scale from 1 to 10), as well as the meaning of the implemented movement (open question: "describe, in your own words, what is the hand doing). Overall, 61.49% of the final stimuli set was accurately described (i.e., open description mentioning grasping, shaking, or absence of movement in the hands' pictures), and the rates of implied motion assessed by a t-test revealed no difference between Grasping and Shaking images ($t(29) = 1.625$, $p = 0.12$) while the implied motion was lower in Still hands than both Grasping hands ($t(29) = 9.373$, $p < 0.001$) and Shaking hands ($t(29) = 11.144$, $p < 0.001$). In the second survey, we aimed at further validating the selected stimuli. We asked another 30 participants to answer two questions for each stimulus. First, they had to complete the sentence "the action performed by the hand is directed towards: …" and chose betwee "an object", "another person", or "neither of the two". Grasping hands were correctly identified as moving towards an object in 90.83% of the cases, shaking hands were correctely identifed as moving towards another person in 90.00% of the cases, while Still hands seemed to be perceived as more ambiguous, with 53.33% of the participants correctly answering "neither of the two". Second, we explicitly asked whether the hand displayed was being still, grasping an object or shaking another hand. Participants correctly answered that our Grasping hands stimuli were grasping an object (95%), that our Social hands were shaking another hand (84.16%) and that our Still hands were being motionless (83.33%).

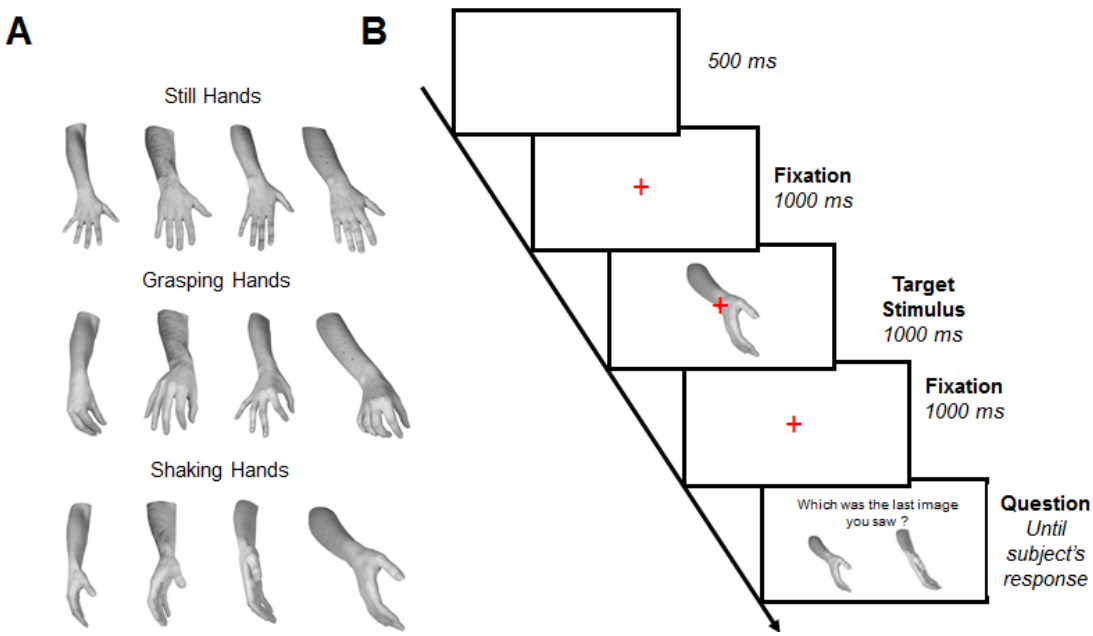


Figure 1| A) Set of stimuli for each of the three classes. **B**) Experimental timeline.

Procedure

Each trial consisted of the presentation of a red fixation cross for 1000 ms before the appearance of the target stimulus. Target images were shown for 500 ms at the center of the screen. The fixation cross remained on the screen during the presentation of the target stimulus in order to avoid EEG artifacts associated with its appearance and disappearance. Stimuli were viewed at a distance of approximately 60 cm. In order to maintain subjects' attention on the stimuli, after the disappearance of the stimulus, in 30% of the trials, participants were asked to choose one among two presented images by pressing one of the two keys from the pad with their right hand to indicate which stimulus they thought was the one presented in that trial (Figure 1B). Stimuli presentation was displayed by using E-Prime 2.0 and was randomized across subjects. The experiment consisted of three blocks with 100 trials each, where stimuli were presented randomly. The total number of trials was 300 (100 trials for each of the three different conditions).

EEG-recordings and preprocessing

EEG signals were recorded and amplified using a Neuroscan SynAmps RT amplifiers system (Compumedics Limited, Melbourne, Australia) and acquired from 56 tin scalp electrodes embedded in a fabric cap (Electro-Cap International, Eaton, OH), arranged according to the 10-10 system. The EEG was recorded from the following channels: Fp1, Fpz, Fp2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FC5, FC3, FC1, FCz, FC2, FC4, FC6, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2. Horizontal electro-oculogram (HEOG) was recorded bipolarly from electrodes placed on the outer canthi of each eye and signals from the left earlobe were also recorded. All electrodes were physically referenced to an electrode placed on the right earlobe and were algebraically re-referenced off-line to the

average of both earlobe electrodes. Impedance was kept below 5 KΩ for all electrodes for the whole duration of the experiment, amplifier hardware band-pass filter was 0.01 to 200 Hz and the sampling rate was 1000 Hz. In order to remove the eye blinks or other artifacts, EEG and horizontal electrooculograms were processed in two separate steps. Data were then downsampled at 500 Hz before a blind source separation method was applied, using Independent Component Analysis (ICA; Jung et al., 2000) implemented in the MATLAB toolbox EEGLab (Delorme and Makeig, 2004) to remove from the EEG any components related to eye movements. The remaining preprocessing steps were performed using the FieldTrip routines (Donders Institute, Nijmegen, The Netherlands; Oostenveld et al., 2011). Namely, trials showing amplifier blocking, residual eye blink, or other types of artifacts were excluded from the analysis manually. After artefact rejection, EEG data were band-pass filtered (high pass = 0.5, low-pass = 40 Hz) and segmented into epochs of 1400 ms from -400 ms to +1000 ms around stimulus onset.

ERP analysis

Analyses of P1, N190 and EPN components were performed in MATLAB R2019a using the FieldTrip routines (Donders Institute, Nijmegen, The Netherlands; Oostenveld et al., 2011). Before ERP averaging across subjects for each condition, EEG epochs were baseline corrected from -200 ms to 0 ms before stimulus presentation. The P1 amplitude was then quantified as the mean amplitude in a time window of 110-170 ms post-stimulus onset, the N190 amplitude as the mean amplitude in a time window of 160-230 ms post-stimulus onset, and the EPN as the mean amplitude in a time window of 250-350 ms post-stimulus onset. Each component's mean amplitudes were further analyzed through a one-way within-subject repeated measures ANOVA with CATEGORY as a factor (i.e., Still, Grasping and Shaking hands) in R (R Core Team, 2014) using the ez-package (Lawrence, 2016). Mean amplitude values were extracted over bilateral occipito-temporal clusters of electrodes (i.e., P7/8 and PO7/8, see (Borhani et al., 2015; Moreau et al., 2018; Moreau et al., 2019). The Greenhouse-Geisser correction for non-sphericity was applied when appropriate (Keselman et al., 1980), and post-hoc comparisons were performed using the Bonferroni correction for multiple comparisons.

Multivariate Pattern Classification Analysis

Preprocessing. The EEG data used for the classification were low-pass filtered at 30 Hz (van Driel et al., 2021) and segmented into epochs of 700 ms (from -200 ms to +500 ms around stimulus onset) in order to reduce computation time. Further preprocessing steps as well all the classification analyses were performed using the MVPA-Light MATLAB toolbox (Treder, 2020). Prior to classification, data were normalized using z-scores to center and scale the training data, providing numerical stability (Treder, 2020). Across all classification analyses, if a class had fewer trials than another, we corrected the imbalance by undersampling the over-represented condition (i.e., randomly removing trials). Then, using the preprocessing option mv_preprocess_average_samples, training data (i.e., EEG trials) from the same class (i.e., Hand Category) were randomly split into 8 groups and averaged, so that the single-trial dataset was replaced by 8 averaged trials (i.e., samples). The classifications

were then run using these averaged samples, as this step is known to increase the signal-to-noise ratio (Grootswagers et al., 2017; Smith and Smith, 2019; Treder, 2020).

Multiclass Classification. We first performed two separate multiclass Linear Discriminant Analysis (LDA) using EEG channels (n = 56) as features to classify the EEG activity in the N190 (averaged over the 160-230 ms interval) and EPN (averaged over the 250-350 ms interval) time windows for each of the three classes (Still, Grasping and Shaking). Classification of the averaged samples generated in the preprocessing (8 samples per class) was performed on Still, Grasping, and Shaking classes together using a k-fold cross-validation procedure in which samples (8 samples x 3 classes = 24 total samples) were divided into 4 folds made of 6 samples each (3 training and 1 testing), repeated 10 times so that each sample was either used for training or testing at least once. The accuracy values across testing folds of all repetitions were then averaged to produce the final classifier performance (see Figure 3A) and also presented on a confusion matrix to assess the probability of the classifier to in/correctly assign classes (see Figure 3B).

Classification across time. We then performed 3 separate binary classification analyses in time (-200 to 500 ms), using Still/Grasping, Still/Shaking, and Grasping/Shaking samples. To do so, a linear L2-regularized Logistic Regression classifier was implemented using the MVPA-Light toolbox and the LIBLINEAR (Fan et al., 2008) functions. For linearly separable data, a Logistic Regression classifier finds the optimal separating hyperplane between classes using a logistic function to model the dependent variable. Instead of choosing an arbitrary hyperparameter C that controls the amount of regularization and adds a penalty for each misclassified data point, we used an automated grid search using nested cross-validation to determine the optimal C (Varoquaux et al., 2017). We also ran the same analysis setting C=1, as it is the most used value in neuroimaging litterature (see Supplementary Figure S1A). Just like we did for the multiclass LDA, we used 4-fold cross-validation repeated 10 times on 8 averaged samples per class (for results on all trials without averaging, see Supplementary Figure S1B). We computed the area under the ROC curve (AUC) to assess classifier performance over time in each of the three binary comparisons. AUC was preferred to the accuracy metric as it is more robust for binary classifications (Stewart et al., 2014). Statistical significance (i.e., performance significantly above 0.5 chance level) was assessed through non-parametric cluster-based permutation tests (Maris and Oostenveld, 2007), using 'maxsum' for cluster corrections and Wilcoxon test for significance assessment.

Classification across time and electrodes. Using identical settings for the classification in time, we performed a spatio-temporal search using two neighboring matrices for time points and electrodes, respectively. Thus, in this analysis each electrode/time-point and its direct neighbors acted as features for the classification, resulting in a channels x time points matrix of AUC scores. By plotting this matrix, we then assessed which electrode carried the most weight in the temporal decoding (see Figure 4).

Results

Behavioral results

Overall, subjects responded correctly at 99.326 % when asked to identify the stimulus presented.

Event-Related Potentials

The results of the ERP analysis are shown in Figure 2.

P1 component. The ANOVA on the P1 mean amplitudes revealed no significant difference between conditions [$F(2,44) = 1.41$, $p = 0.25$, $\eta_p^2 = .060$].

N190 component. The ANOVA on the N190 mean amplitudes revealed a main effect of CATEGORY [$F(2,44) = 10.07$, $p < .001$, $\eta_p^2 = .031$] over the bilateral occipito-temporal cluster (P7, PO7, P8 and PO7). Bonferroni corrected post-hoc comparisons revealed a significantly reduced N190 amplitude for Shaking hands (M = -2.91µV, SD = 2.27) compared to Grasping hands (M = -3.47µV, SD = 2.27) [$p < .001$, see Figure 2] and only a reduction in trend compared with Still hands (M = -3.31µV, SD = 2.05) [$p = 0.06$]. There was no significant difference when comparing Still hands with Grasping hands [$p = 0.50$].

EPN component. The ANOVA on the EPN mean amplitudes revealed a main effect of CATEGORY [$F(2,44) = 21.35$, $p < .001$, $\eta_p^2 = .049$] over the bilateral occipito-temporal cluster (P7, PO7, P8 and PO8). Bonferroni corrected post-hoc comparisons revealed a significantly larger EPN amplitude associated with Shaking hands (M = 1.27µV, SD = 1.31) compared to Grasping hands (M = 1.83µV, SD = 1.45) [$p = .001$] and Still hands (M = 2.01µV, SD = 1.46) [$p < .001$, see Figure 2] and no significant difference between Still and Grasping hands [$p = 0.50$].

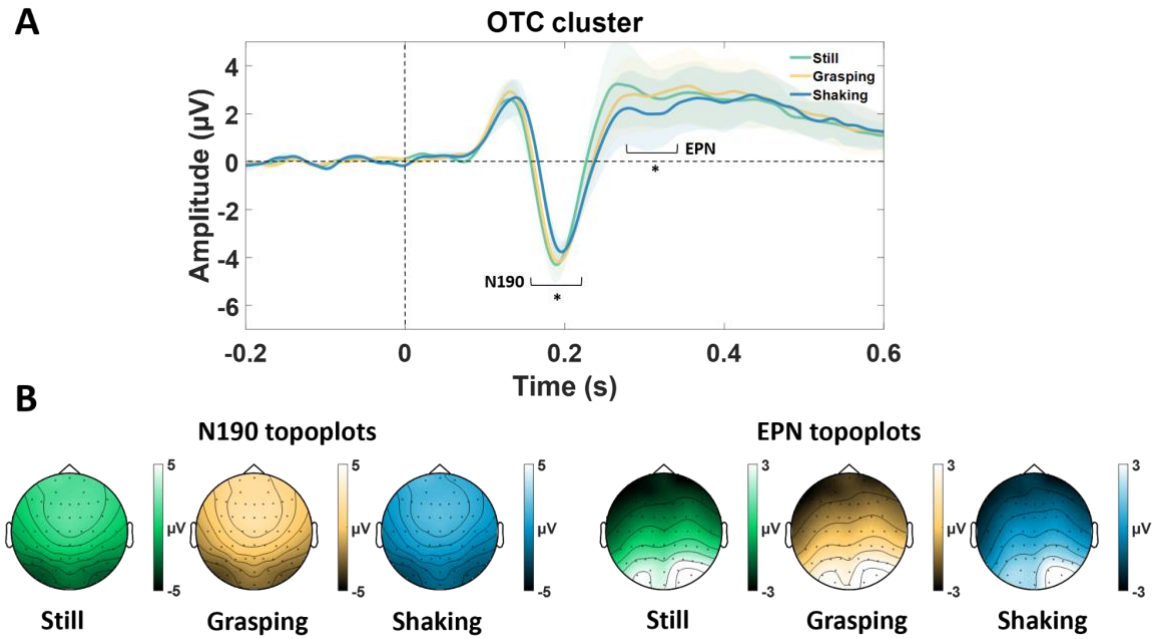


Figure 2| A) N190 and Early Posterior Negativity (EPN) event-related potentials over the bilateral occipito-temporal cluster (P7, PO7, P8, and PO7) for the three conditions; **B**) Topographical plots of the N190 (160-230 ms) and the EPN (250-350 ms) for the three conditions.

Multiclass Classification

Multivariate Pattern Analysis (MVPA) was performed to assess the classification of EEG patterns across the three experimental conditions (i.e., Still, Grasping, and Shaking hands). Our multiclass classification (i.e., Multiclass LDA) successfully discriminated patterns both in the N190 and in the EPN time windows with an accuracy above the chance level of 33% (37% in the N190 time window and 43% in the EPN time window). Interestingly, our results highlight a better overall performance of the classifier in the EPN time window ($t(22) = -3.181$, $p = 0.004$) (Figure 3A).

The confusion matrices, allowing the assessment of the probability for the classifier to incorrectly discriminate between classes of data, also confirm that classification accuracy was above chance for each class (see diagonal from top left to bottom right in the confusion matrices, Figure 3B). However, in the N190 time window, while Still samples were correctly categorized 40% of the time, the classifier shows a high level of confusion between Shaking and Grasping samples (i.e., Shaking hands classified as Grasping in 33% of cases and Grasping hands were categorized as Shaking in 34% of the cases). In the EPN time window, there was an overall higher probability that samples were correctly classified, with Shaking hands being correctly assigned to their true class in 45% of the cases and misclassified as Grasping or Still in only 27% of the cases. These results suggest that EEG patterns in the N190 time-window are more likely to lead to an inaccurate classification across the three categories than patterns in the EPN time window.

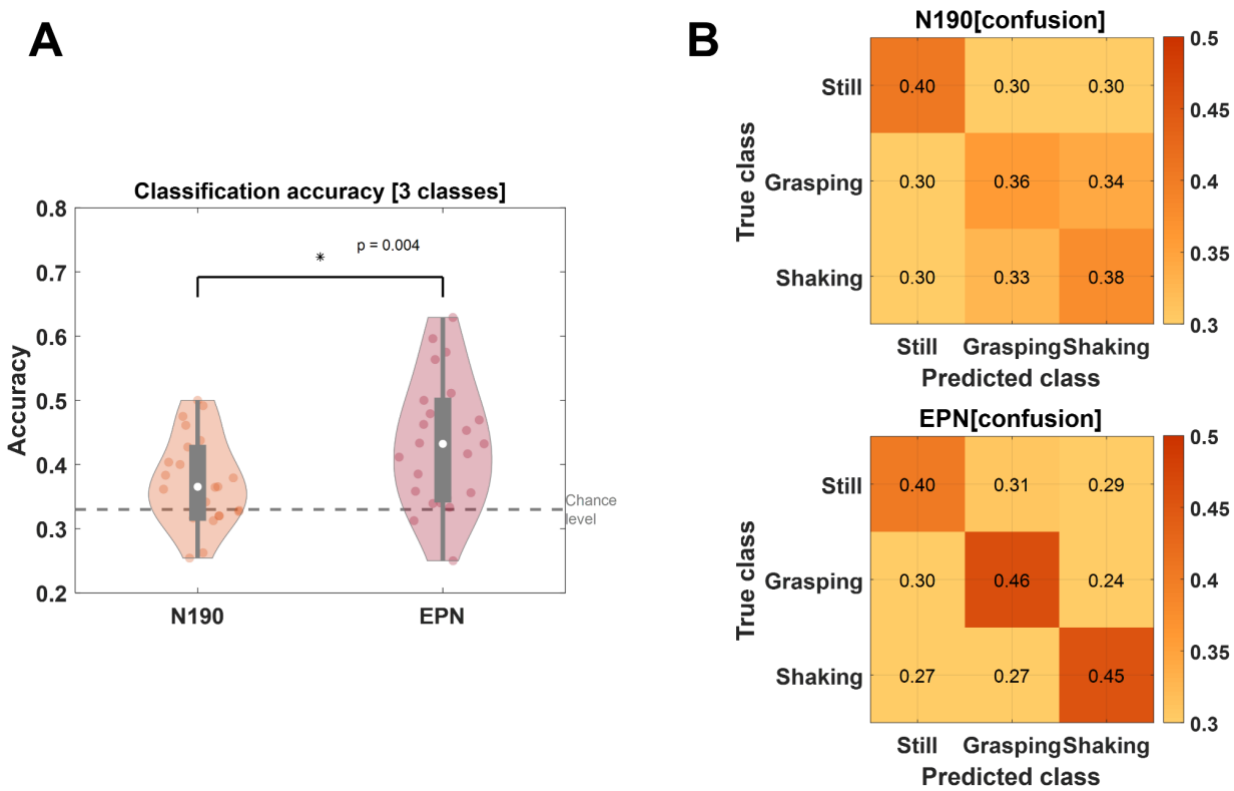
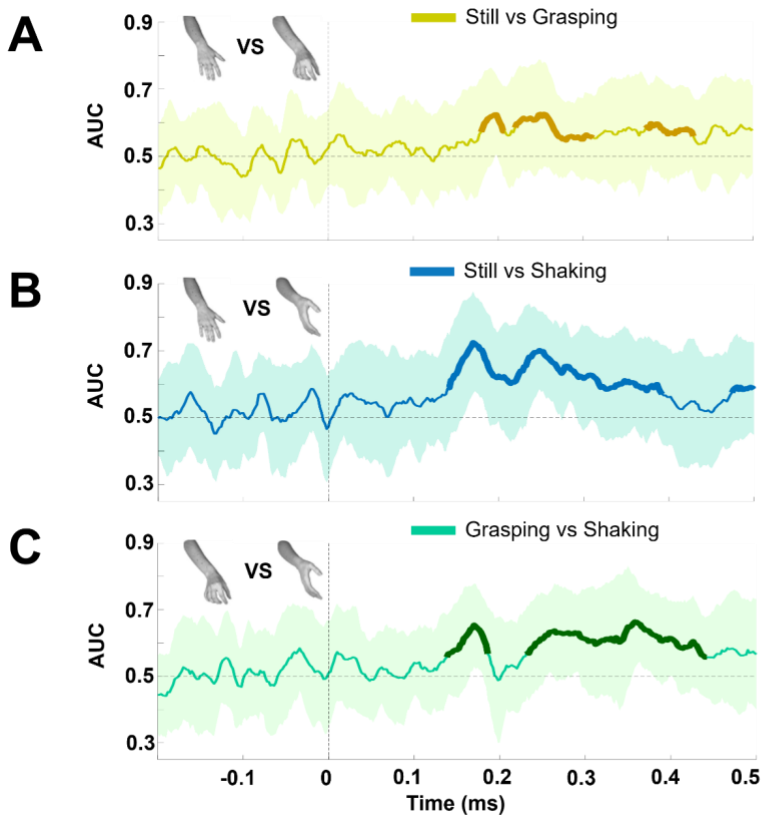


Figure 3| A) Multiclass classification scores (chance level 33%) in the N190 and the EPN time window; **B**) Confusion matrices showing the accuracy of the classifier when in/correctly predicting the class of each sample in the N190 and EPN time window.

Classification across time

To test whether motion and social information displayed in hand stimuli were differently reflected in the EEG time series, we trained and tested a classifier on the EEG data associated with the presentation of those images. Our linear classifier was trained to discriminate between A) Still and Grasping hands, B) Still and Shaking hands, and C) Grasping and Shaking hands (Figure 4 A-C). The results from this analysis plotted as Area Under the ROC Curve (AUC), are shown on the left panels of Figure 4. Statistically significant results ($p < .05$, classification chance level 0.5) are highlighted by a bold line, and the timing and the *p* values of the significant clusters are reported in the tables on the right of Figure 4. First, all classifications show at least one significant cluster, meaning that our classifier was able to discriminate between conditions above the chance level. The Still vs Grasping classification shows the lowest performance values (Figure 4A) and the classifier reaches significance later than the other comparisons (i.e., from 180 ms compared to 140 for B and C). The highest AUC values are observed for the Still vs Shaking classification (i.e., 0.72 at 170ms, see Figure 4B). Crucially, in the Still vs Shaking (Figure 4B) and Grasping vs Shaking (Figure 4C) comparisons, the classifier performance index shows a similar pattern of results, with early significant decoding (starting around 140 ms) and a later robust, significantly above chance performance (until ~400ms). Overall, the classifier is more performant when having to discriminate between Shaking hands from other conditions.



Cluster timing	*p* value
180 – 210 ms	0.033
220 – 310 ms	< 0.001
375 – 430 ms	0.005

Cluster timing	*p* value
140 – 390 ms	< 0.001

Cluster timing	*p* value
140 – 190 ms	0.006
230 – 440 ms	< 0.001

Figure 4| Temporal decoding and significant clusters for the binary comparison of **A**) Still vs Grasping; **B**) Still vs Shaking and, **C**) Grasping vs Shaking.

Classification across time and space

We performed the same binary comparisons used for the above time classification analysis over time and electrodes (adopting a “saliency map” approach, as reported by Vahid et al., (2020) (Figure 5) with the same parameters and statistical analysis (i.e., L2-regularized Logistic Regression classifier, cluster-based permutation test correcting for multiple comparisons). This additional analysis highlights the weight of each electrode in the temporal classification (i.e., how much each electrode contributed to the classifier’s performance). The results can be seen in Figure 5, where only the significantly above chance level ($p < .05$, classification chance level 0.5) points are displayed. Overall, the time and space decoding are in line with the temporal results but reveal two different spatial patterns. We observe a distributed activation pattern across the scalp (i.e., occipital, centroparietal, frontocentral, Figure 5A and 5B) when distinguishing between Still hands and Grasping and Shaking hands (both conveying implied motion), while the decoding between Grasping and Shaking only seem to rely on parieto-occipital sites (matched for implied motion; Figure 5C).

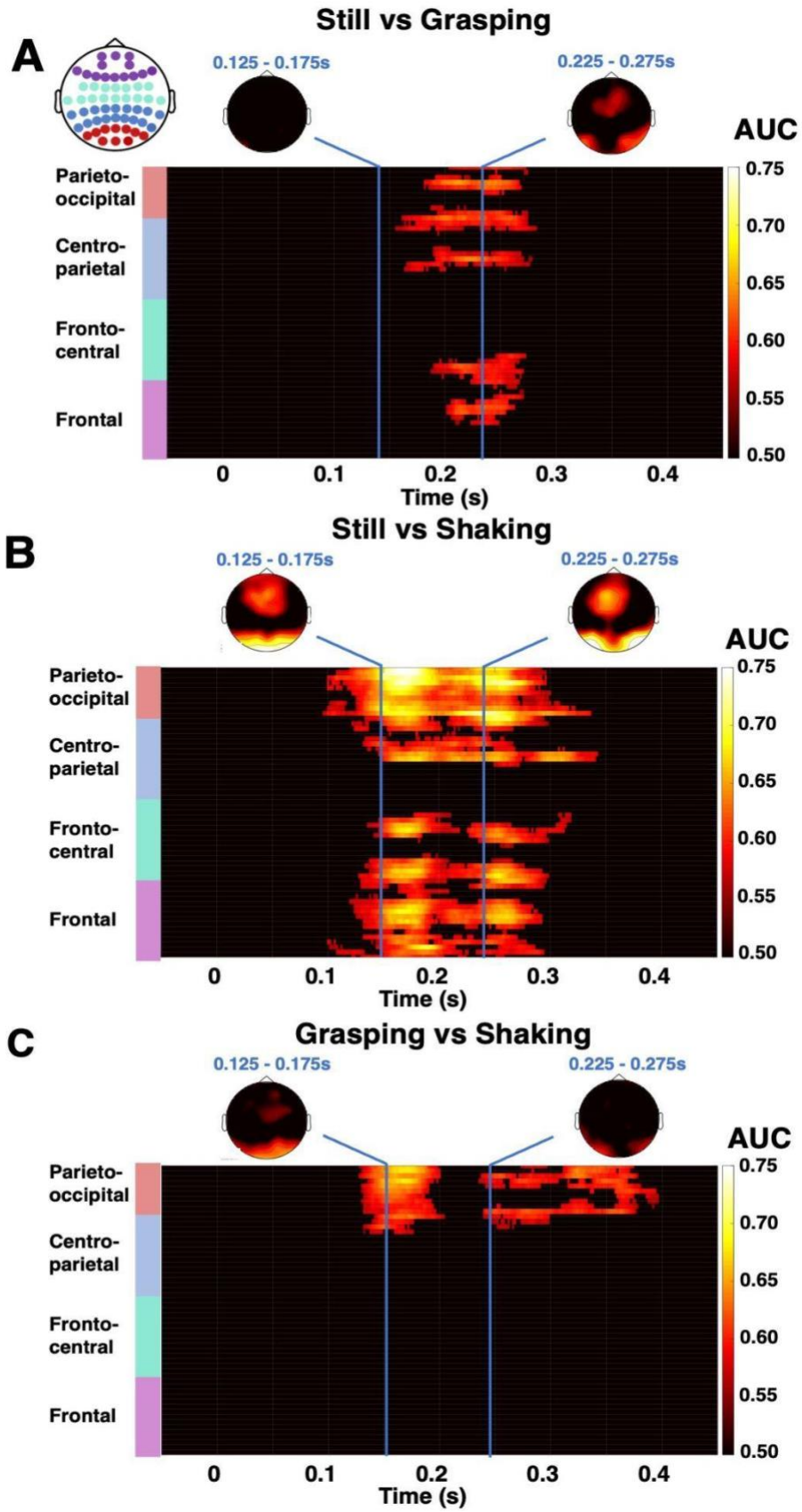


Figure 5| Spatio-temporal decoding and significant clusters for the binary comparison of **A**) Still vs Grasping; **B**) Still vs Shaking and, **C**) Grasping vs Shaking.

Discussion

The aim of the present study was to investigate the EEG markers associated with the visual perception of hand gestures conveying social affordances (shaking hands). By using control stimuli showing hands performing non-social actions (i.e., grasping actions, matched for implied motion with shaking hands) and still hands, we combined univariate analyses of early body-selective N190 (Thierry et al., 2006) and mid-latency attentional selection EPN (Borhani et al., 2015) indexes with multivariate analyses to give a detailed description of the neural (temporal and spatial) dynamics at play during the processing of hand images containing social information. Describing the spatio-temporal dynamics of social affordances processing is important to contribute to the current understanding of the neural underpinnings of social vision (Lingnau & Downing, 2015; Tucciarelli et al., 2019; Tarhan et al., 2020).

ERPs modulations in response to hand gestures conveying social affordances

Our univariate analysis reveals that the early body-selective ERP component (N190) is significantly influenced by the social meaning of the movement implied in hand postures. The N190 component has traditionally been associated with the encoding of structural properties and the categorization of visual inputs as bodies and body parts (Thierry et al., 2006; Urgesi et al., 2007). Interestingly, our findings show a reduction in the amplitude of the N190 for shaking hands compared to the two non-social hands categories (i.e., Grasping and Still, although only marginally significant when comparing Shaking with Still hands). Hence, it appears that the N190 does not only account for the extraction of morphological features of body parts (Taylor et al., 2007; Taylor and Downing, 2011), but is also sensitive to their posture and the meaning of the implied actions. Although a causal contribution of EBA activity to the discrimination of implied hand actions has been excluded by some lesion and brain stimulation studies (Urgesi et al., 2007; Candidi et al., 2008; Moro et al., 2008, but see also, Astafiev et al., 2004 and Lingnau and Downing, 2015), the sensitivity of the N190 (and consequently EBA, its anatomical source, Thierry et al., 2006) to the emotional content of bodily stimuli has been suggested by previous studies (Borhani et al., 2015). Here, we extend these findings and show that hand postures implying a social action differently modulate N190 amplitude compared to hands without social meaning, suggesting an early (200 ms) engagement of the occipito-temporal cortex in the encoding of the sociality of actions (Wurm et al., 2017; Wurm and Caramazza, 2019).

Similarly, we show that the Early Posterior Negativity (EPN) is also differently influenced by the three hand-posture categories, with Shaking hands associated with the largest negative amplitude compared to the two other classes of stimuli. Previous EEG studies suggested that the EPN component indicates early perceptual tagging and prioritized visual processing (Olofsson et al., 2008), hence indexing natural selective attention for arousing and salient stimuli. Others highlighted how this component is strongly affected by negatively affective picture viewing (i.e., distressful scenes, threatening faces, gestures with negative valence, Junghöfer et al., 2001; Schupp, Öhman, et al., 2004; Schupp, Junghöfer, et al., 2004; Flaisch et al., 2009; Flaisch et al., 2011), compared to neutral images. These findings have been framed within the perspective of understanding emotions as action sets (Lang et al., 2013), such that negative visual cues (i.e., insulting gestures, negative facial expressions) elicit preparation for an immediate action. Coherently, the source of the EPN has been located in

the left superior parietal cortex (Jaspers-Fayer et al., 2012), which is also known to mediate visuo-motor transformations (Iacoboni, 2006; Mutha et al., 2011) and is fundamental to mediate interpersonal motor coordination (Sacheli et al., 2015; Era et al., 2018; Era et al., 2020). Here, by showing that the amplitude of the EPN component is sensitive to social content without negative valence, we suggest that the capture of visual attention is not limited to negatively emotional stimuli, but rather to a more general form of preparation that drives attentional states, such as socially salient gestures.

Occipito-parietal electrodes contribute to the decoding of hand gestures conveying social affordances

The univariate results on the modulation of the N190 and EPN by the social affordance conveyed by hand postures are further supported by the performance scores obtained by our multiclass classifier. Specifically, our multiclass LDA classifier successfully discriminated EEG data related to each experimental condition (i.e., Still, Grasping, or Shaking hands) both in the N190 (160-230 ms) and EPN (250-350 ms) time windows, with an overall better performance related to the later one (see Figure 3).

Furthermore, our temporal decoding results suggest that salient information conveyed by hands is encoded in the brain at two different stages. Indeed, our temporal decoding points towards two peaks in decoding accuracy, with an early stage starting before 200ms after stimulus presentation, and a later mid-latency stage in the 250-500ms range. Interestingly, early processing (<200 ms), usually linked to category-selective visual areas, appears to be sensitive to both implied motion and social information (i.e., early accurate classification for Grasping and Shaking hands compared to Still). The temporal decoding analysis is coherent with the ERP analysis in highlighting two key time-sensitive phenomena associated with the visual presentation of hands.

Additionally, occipito-temporal areas are known to encode visual information from socially relevant stimuli (Lingnau and Downing, 2015). However, they are also integrated into a wider fronto-parietal network involved in action understanding and execution, where the activation of frontal and motor areas is thought to facilitate visual perception via top-down connections (Kilner and Frith, 2007; Bressler et al., 2008). Here, as revealed by the classification over time and electrodes (Figure 5), the decoding of hand postures with implied motion (i.e., Shaking and Grasping hands) against Still hands shows that, in addition to the occipito-temporal sites, the performance of the classifier strongly relied on frontal and parietal electrodes. This could suggest top-down modulation processes from motor and parietal areas to the visual processing of hand postures with implied motion compared to Still ones. On the other hand, only occipito-parietal sites significantly contributed to the classifier's performance when differentiating Shaking hands from Grasping hands (i.e., both with implied motion). These results are coherent with previous studies on the role of the parietal cortex in understanding the goal of observed actions in primates (Fogassi et al., 2005; Filippini et al., 2018) and humans (Rizzolatti and Craighero, 2004; Hamilton and Grafton, 2006) and with the suggestion that others' actions may represent social affordances, implying the selection of potential motor responses in parietal cortices (Orban *et al.*, 2021). Furthermore, Wurm and colleagues (2017), using fMRI-multivoxel pattern analysis, have previously reported that sociality of hand actions could be accurately decoded within the lateral occipital cortex (Wurm et al., 2017). Our results

here complement this finding, by showing that the decoding of social and non-social implied actions by hands relies on occipito-parietal and occipito-temporal electrodes and provide new information regarding the precise timings of this process.

Conclusion

To sum up, we combined univariate and multivariate analyses to tackle the different dimensions of the EEG activity associated with the visual processing of hand gestures implying social affordances. This approach allowed us to integrate multiple sources of information, which together provide new evidence on the sensitivity of the N190 and EPN components in discriminating sociality in hand gestures and suggest that the integration of motion and social information happens in the early stages of visual processing, as suggested by our decoding analyses.

References

Argyle, M. (1972). Non-verbal communication in human social interaction. In: *Non-verbal communication*. Oxford, England: Cambridge U. Press, p. xiii, 443–xiii, 443.

Astafiev, S.V., Stanley, C.M., Shulman, G.L., et al. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience*, **7**, 542–48

Avenanti, A., Candidi, M., Urgesi, C. (2013). Vicarious motor activation during action perception: beyond correlational evidence. *Frontiers in Human Neuroscience*, **7**

Borhani, K., Làdavas, E., Maier, M.E., et al. (2015). Emotional and movement-related body postures modulate visual processing. *Social Cognitive and Affective Neuroscience*, **10**, 1092–1101

Bracci, S., Cavina-Pratesi, C., Ietswaart, M., et al. (2012). Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. *Journal of Neurophysiology*, **107**, 1443–56

Bracci, S., Ietswaart, M., Peelen, M.V., et al. (2010). Dissociable Neural Responses to Hands and Non-Hand Body Parts in Human Left Extrastriate Visual Cortex. *Journal of Neurophysiology*, **103**, 3389–97

Bressler, S.L., Tang, W., Sylvester, C.M., et al. (2008). Top-Down Control of Human Visual Cortex by Frontal and Parietal Cortex in Anticipatory Visual Spatial Attention. *The Journal of Neuroscience*, **28**, 10056–61

Campbell, J.I.D., Thompson, V.A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, **44**, 1255–65

Candidi, M., Stienen, B.M.C., Aglioti, S.M., et al. (2011). Event-Related Repetitive Transcranial Magnetic Stimulation of Posterior Superior Temporal Sulcus Improves the Detection of Threatening Postural Changes in Human Bodies. *The Journal of Neuroscience*, **31**, 17547–54

Candidi, M., Stienen, B.M.C., Aglioti, S.M., et al. (2015). Virtual lesion of right posterior superior temporal sulcus modulates conscious visual perception of fearful expressions in faces and bodies. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, **65**, 184–94

Candidi, M., Urgesi, C., Ionta, S., et al. (2008). Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. *Social Neuroscience*, **3**, 388–400

Cattaneo, L., Sandrini, M., Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex (New York, N.Y.: 1991)*, **20**, 2252–58

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, **134**, 9–21

Downing, P.E., Jiang, Y., Shuman, M., et al. (2001). A cortical area selective for visual processing of the human body. *Science (New York, N.Y.)*, **293**, 2470–73

van Driel, J., Olivers, C.N.L., Fahrenfort, J.J. (2021). High-pass filtering artifacts in multivariate classification of neural time series data. *Journal of Neuroscience Methods*, **352**, 109080

Era, V., Aglioti, S.M., Candidi, M. (2020). Inhibitory Theta Burst Stimulation Highlights the Role of Left aIPS and Right TPJ during Complementary and Imitative Human–Avatar Interactions in Cooperative and Competitive Scenarios. *Cerebral Cortex*, **30**, 1677–87

Era, V., Candidi, M., Gandolfo, M., et al. (2018). Inhibition of left anterior intraparietal sulcus shows that mutual adjustment marks dyadic joint-actions in humans. *Social Cognitive and Affective Neuroscience*, **13**, 492–500

Espírito Santo, M.G., Maxim, O.S., Schürmann, M. (2017). N1 responses to images of hands in occipito-temporal event-related potentials. *Neuropsychologia*, **106**, 83–89

Fan, R., Chang, K., Hsieh, C., et al. (2008). LIBLINEAR: A Library for Large Linear Classification

Filippini, M., Breveglieri, R., Hadjidimitrakis, K., et al. (2018). Prediction of Reach Goals in Depth and Direction from the Parietal Cortex. *Cell Reports*, **23**, 725–32

Flaisch, T., Häcker, F., Renner, B., et al. (2011). Emotion and the processing of symbolic gestures: an event-related brain potential study. *Social Cognitive and Affective Neuroscience*, **6**, 109–18

Flaisch, T., Schupp, H.T., Renner, B., et al. (2009). Neural systems of visual attention responding to emotional gestures. *NeuroImage*, **45**, 1339–46

Fogassi, L., Ferrari, P.F., Gesierich, B., et al. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science*, **308**, 662–67

Gandolfo, M., Downing, P.E. (2019). Causal Evidence for Expression of Perceptual Expectations in Category-Selective Extrastriate Regions. *Current Biology*, **29**, 2496-2500.e3

Goodale, M.A. (2014). How (and why) the visual control of action differs from visual perception. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20140337

Grootswagers, T., Wardle, S.G., Carlson, T.A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, **29**, 677–97

Gschwind, M., Pourtois, G., Schwartz, S., et al. (2012). White-Matter Connectivity between Face-Responsive Regions in the Human Brain. *Cerebral Cortex*, **22**, 1564–76

Hamilton, A.F. de C., Grafton, S.T. (2006). Goal Representation in Human Anterior Intraparietal Sulcus. *Journal of Neuroscience*, **26**, 1133–37

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, **51**, 59–67

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, **4**, 223–33

Hugenberg, K., Wilson, J.P. (2013). Faces are central to social cognition. In: *The Oxford handbook of social cognition*. Oxford library of psychology. New York, NY, US: Oxford University Press, p. 167–93.

Iacoboni, M. (2006). Visuo-motor integration and control in the human posterior parietal cortex: evidence from TMS and fMRI. *Neuropsychologia*, **44**, 2691–99

Ishizu, T., Amemiya, K., Yumoto, M., et al. (2010). Magnetoencephalographic study of the neural responses in body perception. *Neuroscience Letters*, **481**, 36–40

Jaspers-Fayer, F., Ertl, M., Leicht, G., et al. (2012). Single-trial EEG–fMRI coupling of the emotional auditory early posterior negativity. *NeuroImage*, **62**, 1807–14

Jung, T.P., Makeig, S., Humphries, C., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, **37**, 163–78

Junghöfer, M., Bradley, M.M., Elbert, T.R., et al. (2001). Fleeting images: A new look at early emotion discrimination. *Psychophysiology*, **38**, 175–78

Keselman, H.J., Rogan, J.C., Mendoza, J.L., et al. (1980). Testing the validity conditions of repeated measures F tests. *Psychological Bulletin*, **87**, 479–81

Kilner, J.M., Frith, C.D. (2007). A possible role for primary motor cortex during action observation. *Proceedings of the National Academy of Sciences*, **104**, 8683–84

Lang, P.J., Simons, R.F., Balaban, M., et al. (2013). *Attention and Orienting: Sensory and Motivational Processes*. Psychology Press.

Lawrence, M. ez: Easy analysis and visualization of factorial experiments [online]. Available from: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=7gTDUWMAAAAJ&citation_for_view=7gTDUWMAAAAJ:WF5omc3nYNoC [Accessed July 15, 2021].

Lingnau, A., Downing, P.E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, **19**, 268–77

Maris, E., Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, **164**, 177–90

Meeren, H.K.M., Gelder, B. de, Ahlfors, S.P., et al. (2013). Different Cortical Dynamics in Face and Body Perception: An MEG study. *PLOS ONE*, **8**, e71408

Milner, A.D., Goodale, M.A. (2008). Two visual systems re-viewed. *Neuropsychologia*, **46**, 774–85

Milner, D., Goodale, M. (2006). *The Visual Brain in Action*. OUP Oxford.

Moreau, Q., Parrotta, E., Era, V., et al. (2019). Role of the occipito-temporal theta rhythm in hand visual identification. *Journal of Neurophysiology*, **123**, 167–77

Moreau, Q., Pavone, E.F., Aglioti, S.M., et al. (2018). Theta synchronization over occipito-temporal cortices during visual perception of body parts. *European Journal of Neuroscience*, **48**, 2826–35

Moro, V., Pernigo, S., Avesani, R., et al. (2012). Visual body recognition in a prosopagnosic patient. *Neuropsychologia*, **50**, 104–17

Moro, V., Urgesi, C., Pernigo, S., et al. (2008). The Neural Basis of Body Form and Body Action Agnosia. *Neuron*, **60**, 235–46

Mutha, P.K., Sainburg, R.L., Haaland, K.Y. (2011). Left Parietal Regions Are Critical for Adaptive Visuomotor Control. *The Journal of Neuroscience*, **31**, 6972–81

Olofsson, J.K., Nordin, S., Sequeira, H., et al. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, **77**, 247–65

Oostenveld, R., Fries, P., Maris, E., et al. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, **2011**, 156869

Orban, G.A., Lanzilotto, M., Bonini, L. (2021). From Observed Action Identity to Social Affordances. *Trends in Cognitive Sciences*, **25**, 493–505

Orlov, T., Makin, T.R., Zohary, E. (2010). Topographic Representation of the Human Body in the Occipitotemporal Cortex. *Neuron*, **68**, 586–600

Pitcher, D., Ungerleider, L.G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, **25**, 100–110

Ramsey, R. (2018). Neural Integration in Body Perception. *Journal of Cognitive Neuroscience*, **30**, 1442–51

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

Sacheli, L.M., Candidi, M., Era, V., et al. (2015). Causative role of left aIPS in coding shared goals during human–avatar complementary joint actions. *Nature Communications*, **6**, 7544

Sadeh, B., Pitcher, D., Brandman, T., et al. (2011). Stimulation of category-selective brain areas modulates ERP to their preferred categories. *Current biology: CB*, **21**, 1894–99

Sato, W., Kochiyama, T., Yoshikawa, S., et al. (2001). Emotional expression boosts early visual processing of the face: ERP recording and its decomposition by independent component analysis. *Neuroreport*, **12**, 709–14

Schupp, H.T., Junghöfer, M., Weike, A.I., et al. (2004). The selective processing of briefly presented affective pictures: An ERP analysis. *Psychophysiology*, **41**, 441–49

Schupp, H.T., Öhman, A., Junghöfer, M., et al. (2004). The Facilitated Processing of Threatening Faces: An ERP Analysis. *Emotion*, **4**, 189–200

Sliwinska, M.W., Bearpark, C., Corkhill, J., et al. (2020). Dissociable pathways for

moving and static face perception begin in early visual cortex: Evidence from an acquired prosopagnosic. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, **130**, 327–39

Smith, F.W., Smith, M.L. (2019). Decoding the dynamic representation of facial expressions of emotion in explicit and incidental tasks. *NeuroImage*, **195**, 261–71

Stewart, A.X., Nuthmann, A., Sanguinetti, G. (2014). Single-trial classification of EEG in a visual object task using ICA and machine learning. *Journal of Neuroscience Methods*, **228**, 1–14

Tarhan, L., Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, **11**, 3002

Taylor, J.C., Downing, P.E. (2011). Division of Labor between Lateral and Ventral Extrastriate Representations of Faces, Bodies, and Objects. *Journal of Cognitive Neuroscience*, **23**, 4122–37

Taylor, J.C., Wiggett, A.J., Downing, P.E. (2007). Functional MRI Analysis of Body and Body Part Representations in the Extrastriate and Fusiform Body Areas. *Journal of Neurophysiology*, **98**, 1626–33

Thierry, G., Pegna, A.J., Dodds, C., et al. (2006). An event-related potential component sensitive to images of the human body. *NeuroImage*, **32**, 871–79

Treder, M.S. (2020). MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Frontiers in Neuroscience*, **0**

Tucciarelli, R., Turella, L., Oosterhof, N.N., et al. (2015). MEG Multivariate Analysis Reveals Early Abstract Action Representations in the Lateral Occipitotemporal Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **35**, 16034–45

Tucciarelli, R., Wurm, M., Baccolo, E., et al. (2019). The representational space of observed actions M. Lescroart, R. B. Ivry, M. Lescroart, et al. (eds). *eLife*, **8**, e47686

Urgesi, C., Berlucchi, G., Aglioti, S.M. (2004). Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Current biology: CB*, **14**, 2130–34

Urgesi, C., Calvo-Merino, B., Haggard, P., et al. (2007). Transcranial Magnetic Stimulation Reveals Two Cortical Pathways for Visual Body Processing. *Journal of Neuroscience*, **27**, 8023–30

Urgesi, C., Candidi, M., Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in Human Neuroscience*, **8**

Vahid, A., Mückschel, M., Stober, S., et al. (2020). Applying deep learning to single-trial EEG data provides evidence for complementary theories on action control. *Communications Biology*, **3**, 1–11

Varoquaux, G., Raamana, P.R., Engemann, D.A., et al. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, **145**, 166–79

Wurm, M.F., Caramazza, A. (2019). Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract representations of sociality. *NeuroImage*, **202**, 116153

Wurm, M.F., Caramazza, A., Lingnau, A. (2017). Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. *The Journal of Neuroscience*, **37**, 562–75

SUPPLEMENTARY MATERIALS

Classification across time

In order to validate our multivariate results we run different classification analyses across time, varying the hyperparameters assigned to the classifier. Specifically, first, we checked whether the results would be affected by removing the grid search procedure we adopted to find the optimal regularization parameter C, and setting it to 1. The results from this analysis, plotted as Area Under the ROC Curve (AUC), are shown on Figure S1a. Statistically significant results ($p < .05$, classification chance level 0.5) are highlighted by a bold line. The results match the ones obtained finding the best C using grid search, with the highest AUC values observed for the Still vs Shaking classification and the classifier performance showing a similar pattern of results in the Still vs Shaking and Grasping vs Shaking comparisons, with an early significant decoding (starting around 140 ms) and a later robust, significantly above chance performance (until around 400ms).

Second, in order to check whether our preprocessing step of averaging 8 samples before running the k-fold classification might have caused an overfitted performance, we re-run our analysis reducing the averaged samples to 4. The results from this analysis, plotted as Area Under the ROC Curve (AUC), are shown on Figure S1b. Statistically significant results ($p < .05$, classification chance level 0.5) are highlighted by a bold line. The results match the ones obtained averaging 8 samples before the classification, with a general reduction in the performance scores of the classifier.

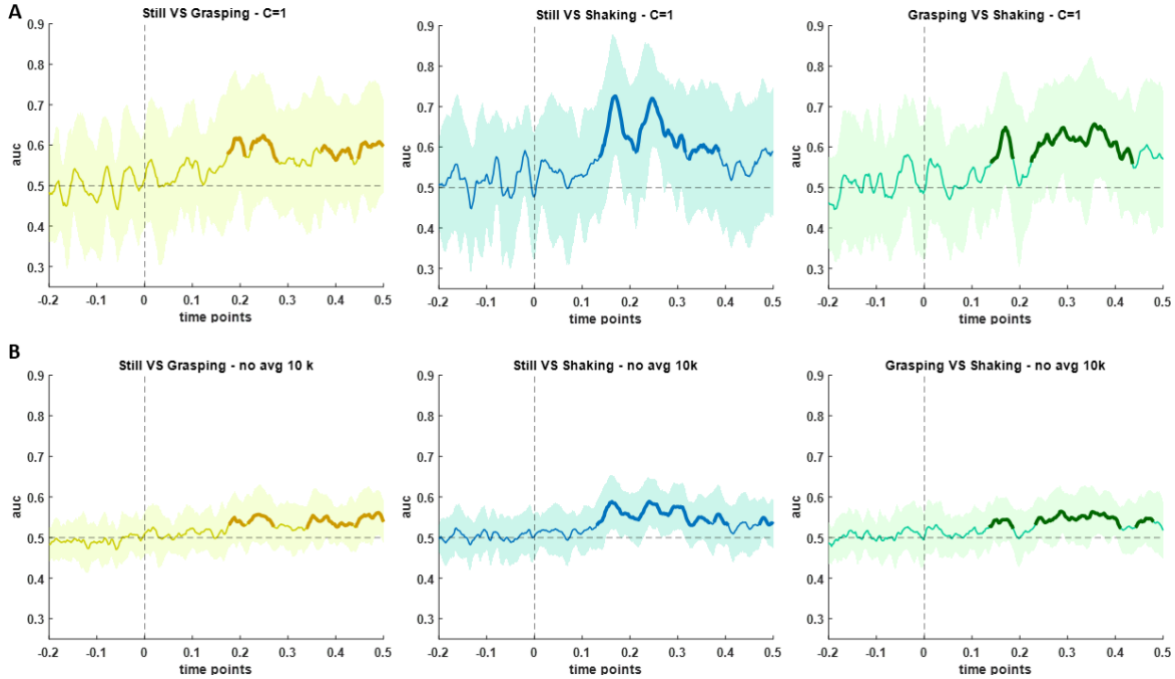


Figure S1| Temporal decoding and significant clusters for each binary classification. **A)** results obtained with the same parameters used in the main analysis (L2 Logistic Regression classifier, averaging 8 sample, 4-fold cross-validation repeated 10 times) except the setting of the hyperparameter C = 1 instead of using a grid search to find its optimal value. **B)** results obtained with the same parameters used in the main analysis (L2 Logistic Regression classifier, hyperparameter C found using grid search, 4-fold cross-validation repeated 10 times) except the preprocessing step of averaging 4 samples instead of 8 in order to increase the SNR.