Robust memory of face moral values is encoded in the human caudate tail: A simultaneous EEG-fMRI study

Ali Ataei [a,b], Arash Amini [a], Ali Ghazizadeh [a,b,c]

[a] EE Department, Sharif university of technology, Tehran, Iran

[b] Sharif Brain Center, Sharif university of technology, Tehran, Iran

[c] School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran

Corresponding author: Ali Ghazizadeh, EE Department, Sharif University of Technology, Tehran, Azadi Avenue, postal code:1458889694 , ghazizadeh@sharif.edu

Abstract:

Moral judgements about people based on their actions is a key component that guides social decision making. It is currently unknown how positive or negative moral judgments associated with a person's face are processed and stored in the brain for a long time. Here, we investigate the long-term memory of moral values associated with human faces using simultaneous EEG-fMRI data acquisition. Results show that only a few exposures to morally charged stories of people are enough to form long-term memories a day later for a relatively large number of new faces. Event related potentials (ERPs) showed a significant differentiation of remembered good vs bad faces over centerofrontal electrode sites (value ERP). EEG-informed fMRI analysis revealed a subcortical cluster centered on the left caudate tail (CDt) as a correlate of the face value ERP. Importantly neither this analysis nor a conventional whole brain analysis revealed any significant coding of face values in cortical areas, in particular the fusiform face area (FFA). Conversely an fMRI-informed EEG source localization using accurate subject-specific EEG head models also revealed activation in the left caudate tail. Nevertheless, the detected caudate tail region was found to be functionally connected to the FFA, suggesting FFA to be the source of face-specific information to CDt. These results identify CDt as a main site for encoding the long-term value memories of faces in humans suggesting that moral value of faces activates the same subcortical basal ganglia circuitry involved in processing reward value memory for objects in primates.

Introduction:

Our past experiences with people whether good or bad affect our future interactions with them and guide our social decision making. In particular, moral judgement about people based on their actions is a key component in our evaluation of an individual (Cornwell & Higgins, 2019; Jiang

et al., 2022)(Greene, 2009). It is currently unknown how positive or negative moral judgments associated with a person's face are processed and stored in the brain in long-term memory.

Human brain has specialized face processing areas in the temporal cortex including the occipital face area (OFA), superior temporal sulcus (STS) and the fusiform face area (FFA) (Duchaine & Yovel, 2015; Kanwisher & Yovel, 2006). Several studies have examined the coding of various aspects of faces based on their intrinsic visual features such as emotional expressions (Engell & Haxby, 2007; Winston et al., 2004), attractiveness (Cloutier et al., 2008; Kranz & Ishai, 2006; O'Doherty et al., 2003; Said et al., 2011; Winston et al., 2007) trustworthiness (Engell et al., 2007; Winston et al., 2002) or social value (Oosterhof & Todorov, 2008; Todorov et al., 2011) in the human brain with somewhat divergent results. A review and meta-analysis of these findings suggests consistent activations for negative evaluations in the amygdala, and for positive evaluations in the medial orbitofrontal cortex (mOFC), anterior cingulate cortex (ACC), caudate nucleus and the nucleus accumbens (NAcc) (Mende-Siedlecki et al., 2013).

The value circuitry for objects in general is extensively studied in both humans and non-human primates with key cortical and subcortical areas such as orbitofrontal cortex (OFC), insula, ACC, basal ganglia, amygdala and midbrain dopaminergic areas being activated during object reward association tasks (Berridge & Kringelbach, 2008; Kim et al., 2014; Morrison & Salzman, 2010; Padoa-Schioppa & Assad, 2006; Rushworth & Behrens, 2008; Schultz, 2007) . Value memory is also shown to activate a temporal-prefrontal circuitry along with its functionally connected subcortical areas, in particular, the caudate nucleus, amygdala and claustrum (Ghazizadeh, Griggs, et al., 2018; Ghazizadeh, Hong, et al., 2018; Ghazizadeh & Hikosaka, 2021; Kang et al., 2021; Kim & Hikosaka, 2013; Yasuda et al., 2012). Interestingly, these studies have illuminated some different aspects of the short and long-term value memories. Specifically, it is shown that while the head of caudate nucleus exclusively represents the flexible and short-term memory of object values, the tail of caudate nucleus exclusively represents the stable and long-term memory of object values in monkeys (Kim et al., 2014). Replication of the same experimental task on humans using fMRI (Farmani Sepideh et al., n.d.; Kang et al., 2021) has revealed activities in the ventral striatum, caudate body and hypothalamus. While there are some recent studies that have looked at short-term effects of value assignment to faces using EEG or fMRI (Baum & Abdel Rahman, 2021b, 2021a; Cloutier & Gyurovski, 2014; Luo et al., 2019; Schindler et al., 2021; Singer et al., 2004; Suess et al., 2015; Todorov et al., 2007), it is not known how long-term memory of associated value with faces engages this circuitry and whether it can change even the primary face processing in areas such as FFA.

To address this question, we randomly assigned positive or negative values based on morally charged biographical stories to novel faces and then examined the brain activations to these faces a day later using a simultaneous EEG-fMRI paradigm which involved a binary choice for face values. The value ERP showed a significant differentiation of the correctly identified good and bad faces over center-frontal electrodes, peaking at about 600ms post-stimulus and lasting until almost the end of face presentation. This ERP is shown to originate from the left caudate

tail, based on results from both EEG-informed fMRI analysis and fMRI-informed EEG source localization. Interestingly, while none of the cortical face processing areas were found to encode the moral values of faces, some were found to be functionally connected to the detected value-coding caudate tail region.

Results:

To create value memory for faces, we used 24 arbitrary artificial faces created by StyleGAN2 (Karras et al., 2019) . Each face was randomly associated with a brief unique biography with either a positive (good faces) or a negative (bad faces) moral value (see supplementary table 1 for a list of all stories). In the value training session (Fig. 1a), subjects viewed each face while listening to its short story. Each face was viewed for 10 sec within a block of 24 faces and the process was repeated 2 times, 2 hours apart. The assignment of biographies to the faces were swapped across so that each face was associated with positive values for half of the subjects and with negative values for the other half.

One day later, we tested the face value memory of subjects in the MRI scanner being simultaneously equipped with an MRI-compatible EEG cap (value memory session). During the experimental task, each face was portrayed for 2.5s (Fig. 1b). Then, a black screen was shown with two letters of "G" and "B" (referring to "Good" and "Bad", respectively) on the left or right visual hemifield for 2s, during which the subject had to indicate his/her response by pressing a response button using the corresponding hand. The location of letters "G" and "B" were randomly flipped in each trial. The subjects were instructed to answer all trials based on their closest guess, even if they did not remember an explicit history about a face. The subjects' accuracy in identifying face types was 78% on average, which is significantly higher than the chance level. The subjects' accuracies for each specific category were also significantly high (Fig. 1c; t-test, $p < 1e-6$ & $p < 1e-8$, for bad and good faces respectively). This suggests that a brief exposure to a large number of new people (24 new faces) and their moral stories creates a lasting memory that is accessible a day later. Nevertheless, the performance for the good faces was significantly higher than that for bad ones by about 12% (Fig. 1c; paired t-test, $p = 0.004$). To ensure that there was no systematic bias for some faces to be remembered as good or bad, we looked at the percentage of times a certain face was chosen as good or bad across subjects (including additional subjects outside the EEG-fMRI experiment, n= 34). Results showed that among the 24 faces used, only 2 showed significant bias to be chosen as good (Suppl. Table 3). We conducted our subsequent analysis both by inclusion and exclusion of these 2 faces, which did not change any of our main findings as will be discussed.
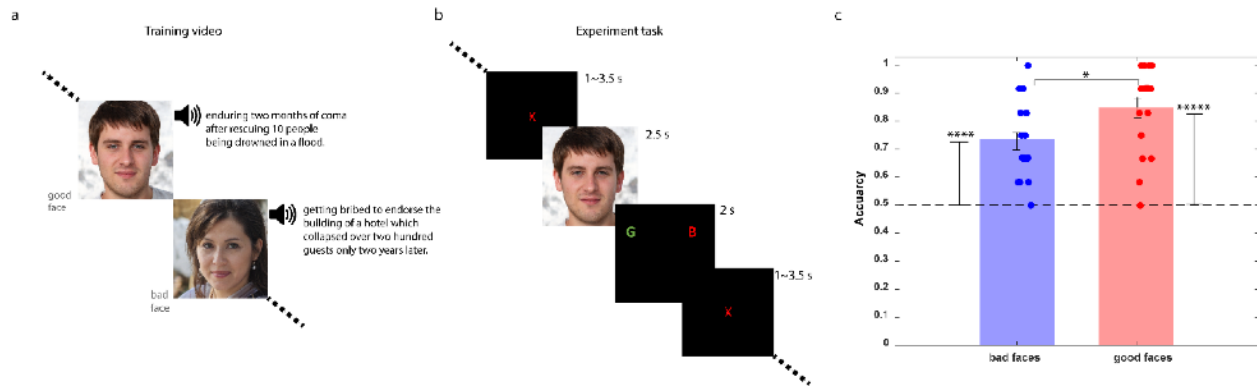
Figure 1: Subjects' training procedure, memory task and behavioral results. a) Value training video: In this video, 24 novel faces were associated with short biographical stories, each conveying a positive or negative moral value about the presented face (good or bad faces, see suppl. Table 1 for the list of morally-charged stories). **b)** Memory task: faces seen the day before were shown in a random sequence for each subject. Each face was shown for 2.5s. Then, a black screen was shown with two letters of "G" and "B" (referring to "good" and "bad", respectively) for 2s during which, the subject had to indicate his/ her response by pressing the appropriate button. The sides of letters "G" and "B" were randomly flipped in each trial. Then, an inter-stimulus-interval (a black screen with a red fixation cross at center) followed for a random time period between 1 to 3.5s. **c)** Behavioral result: The performance of subjects in judging face values, along with the chance level (=50%, dashed line) is shown. Individual subject data points are shown. Error bars indicate standard errors.

Robust differentiation of remembered good and bad faces over center-frontal electrodes

To study the differential neuronal activity during remembered good and bad faces, event-related potentials (ERPs) for each category, separately were calculated. In order to create a more precise group average, the EEG signals for each subject were normalized and transferred to the standard cap situated over the standard brain (see methods for details). The ERPs for both good and bad faces in electrode P8, showed significant negativities around 230ms post stimulus (lasting from about 190ms to 250ms, suppl. Fig. 2 a,b; one-sided t-test against baseline; p-value < 0.05, FDR-corrected) consistent with previously reported N170 and early posterior negativity (EPN) components (Baum & Abdel Rahman, 2021b; Schindler et al., 2021). The normalized z-scored ERPs for the remembered good and bad faces, as well as their difference (the value ERP) showed a robust differentiation between good and bad faces over a large portion of the 2.5s stimulus presentation on center-frontal channels such as C1 and FC1 (Fig. 2a; one-sided t-test against baseline; p-value ≤0.05, Bonferroni corrected over channels and time). This value dissociation is more strongly evident in the interval of 0.5 to 1.5s post-stimulus (raw non-normalized ERPs showed a similar value differentiation with a similar time course, suppl. Fig. 2c).
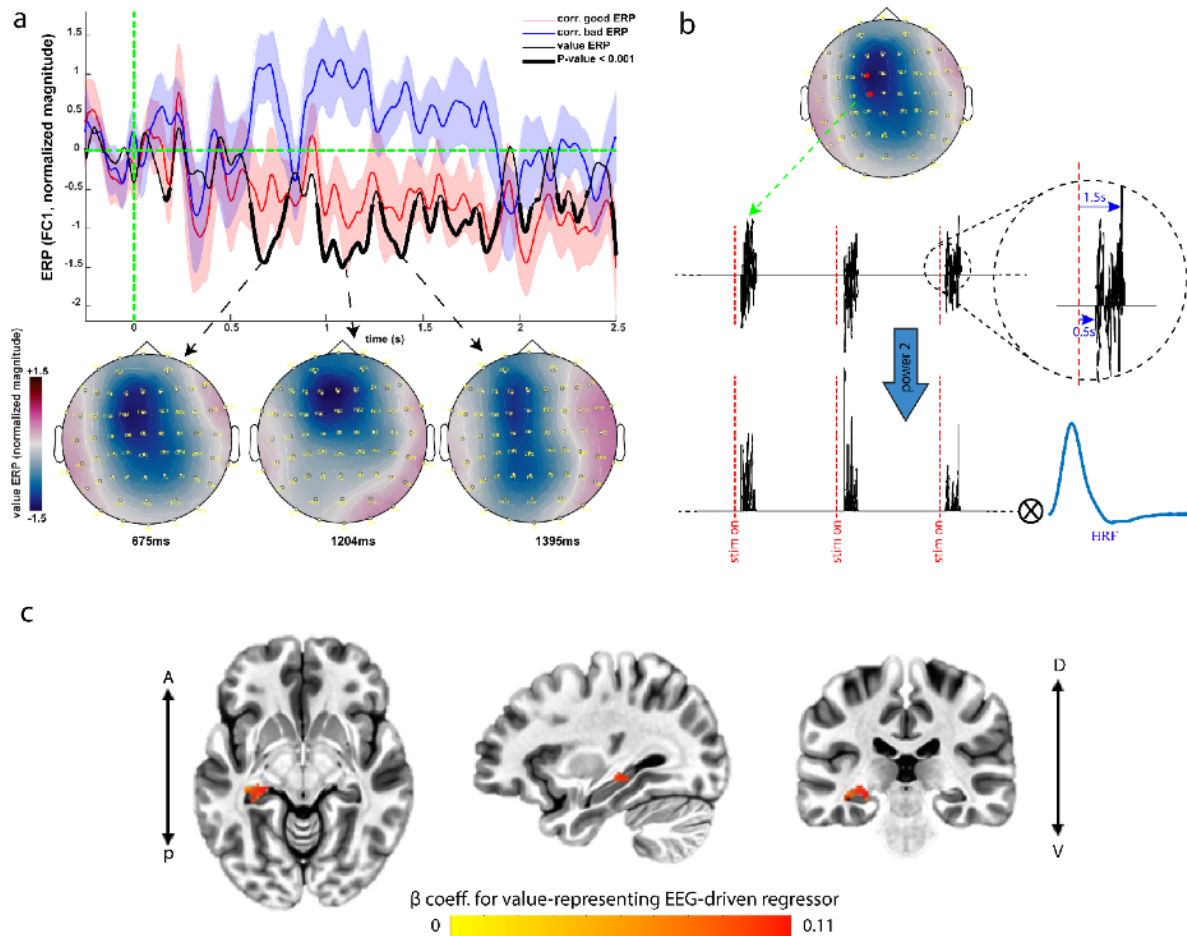
Figure 2: Value ERP & the EEG-informed fMRI analysis. a) The group-average ERPs for the memorized good and bad faces and their difference (value ERP) in electrode FC1 along with the associated group-average scalp potential topographies at some sample time points. Parts of the value ERP that are significantly different from its baseline are shown with bolded black segments (p-value < 0.05, Bonferroni-corrected over channels and time), **b)** The average of electric potentials in center-frontal electrodes over the significant interval (0.5-1.5s post-stimulus) is considered as the EEG-driven value signal. This signal is then squared and consequently convolved with the canonical HRF to build our EEG-driven regressor. **c)** EEG-informed fMRI analysis: the group-average of the beta coefficients for the EEG-driven regressor shows significant activation in a cluster centered on the left caudate tail (p-value < 0.01, cluster-corrected).

EEG-informed fMRI analysis reveals caudate tail as the origin of the value ERP

We next performed an EEG-informed fMRI analysis. Based on the time course of value ERP, we used the average EEG signal over C1 and FC1 in the interval from 0.5s to 1.5s post-stimulus as our EEG-driven regressor in the fMRI GLM model (Fig.2b, see methods). We have previously shown that since the BOLD signal measures the energy consumption in the brain, the square (or power 2) of the EEG signals is the optimal regressor of the BOLD responses (Ataei et al., 2022).

Thus, to localize the source(s) of EEG signals in the center-frontal electrodes, its square was used as the *trial by trial* correlate of face value signal in the brain (Fig.2b).

Other regressors in the model included correct answer to good faces, "correct good", correct answer to bad faces, "correct bad", incorrect answer to good faces, "incorrect good" and incorrect answer to bad faces, "incorrect bad". In case of no incorrect answer to a certain category, the associated regressor was omitted for that subject. Moreover, to account for the effect of subject's motor action during his/her button press response, we also added two confounding regressors, one for each hand. Finally, the average BOLD signal inside the ventricles was used as another nuisance regressor to ensure that responses in structures near the ventricles are not affected by such nonneural extraneous signals (GLM1, see materials & methods for details of the regressor design). As a control analysis, the right versus left hand contrast showed significant activation in the contralateral and deactivation in the ipsilateral motor cortices (suppl. Fig. 3a, p-value < 0.01, cluster-corrected). Moreover, activation of the ipsilateral and deactivation of the contralateral cerebellar cortices were also observable (suppl. Fig. 3b, p-value < 0.01, cluster-corrected), consistent with existing literature (Thickbroom et al., 2003).

The BOLD correlate of the EEG-driven regressor showed significant activation only in a subcortical cluster centered on the left caudate tail (CDt) (Fig. 2c, p-value < 0.01, cluster-corrected). Notably, the contrast of correct good vs. correct bad (the value contrast) did not reveal any significant activation in the brain suggesting that the source of observed EEG responses in the center-frontal electrodes were most likely limited to CDt and that this activity was well captured by the EEG power of center-frontal electrodes in each trial. This is expected since a correct model of statistical dependence among variables predicts that given the EEG signal originating from value contrast, BOLD should become independent of value contrast itself (suppl. Fig. 4a).

To ensure that CDt is indeed activated by the value contrast, we repeated our GLM analysis without an EEG-driven regressor (GLM2). As expected, in this case correct good vs. correct bad contrast showed significant activation in the left CDt (suppl. Fig.3b). We note however that the extent of activation in this case was broader (including parts of the posterior hippocampus, table 1) compared to the more localized CDt activation seen in the EEG-informed analysis attesting to the advantage of simultaneous EEG-fMRI paradigm used in this study. Importantly, we observed no other significant subcortical or cortical value differentiations in either GLM1 or GLM2, suggesting a special role for CDt in representing long-term memory of moral value of faces. Repeating the analysis by excluding the two faces that had a choice bias across subjects did not change our results confirming the activations found in the left caudate tail. (suppl. Fig. 5, p-value < 0.01, cluster-corrected).

fMRI-informed EEG source localization confirms left CDt as the origin of the value ERP

While the spatial resolution of EEG is low, the resolution of its localized sources is highly sensitive to the implemented head model and the method used to solve the inverse problem (Michel & Brunet, 2019; Michel & He, 2019). Conversely one may use fMRI-informed EEG source localization to find sources of neural activity of interest. Here we tried this approach by performing a source localization for each subject using a "mixed" forward model (assuming cortical dipoles perpendicular to the cortex surface and subcortical ones with unknown directions, see methods for details) conducted based on subject-specific MRI images. In our mixed model, we included all the neocortex, but selected subcortical structures based on fMRI results, to allow the low-SNR[1] subcortical sources to be detected more robustly. Specifically, we included the caudate and hippocampus as they were found in the whole brain fMRI analysis (GLM2). We also included amygdala both as a subcortical benchmark and for its well-known role in value coding (Wassum & Izquierdo, 2015). Notably, the group average of the EEG sources revealed significant activity in the left caudate tail and deactivation of the left anterior hippocampus (Fig. 3a; $p \leq 0.001$, duration $\geq$ 150ms & cluster > 3 dipoles). Interestingly such deactivation of anterior hippocampus can be seen in whole brain fMRI analysis (GLM2) as well if no cluster correction is used. (suppl. Fig. 6b). No significant activity was found in cortical areas ($p \leq 0.001$, duration $\geq$ 150ms & cluster > 7 dipoles) or in amygdala (duration $\geq$ 150ms & cluster > 3 dipoles). The time course of activity in the left caudate tail (Fig. 3b) shows the emergence of value memory at about 550ms after face presentation, roughly consistent with the onset of the value ERP seen in center-frontal electrodes (Fig. 2a). Also note that the sign of activities for the subcortical regions with unknown dipole directions is ambiguous (see methods for more details).
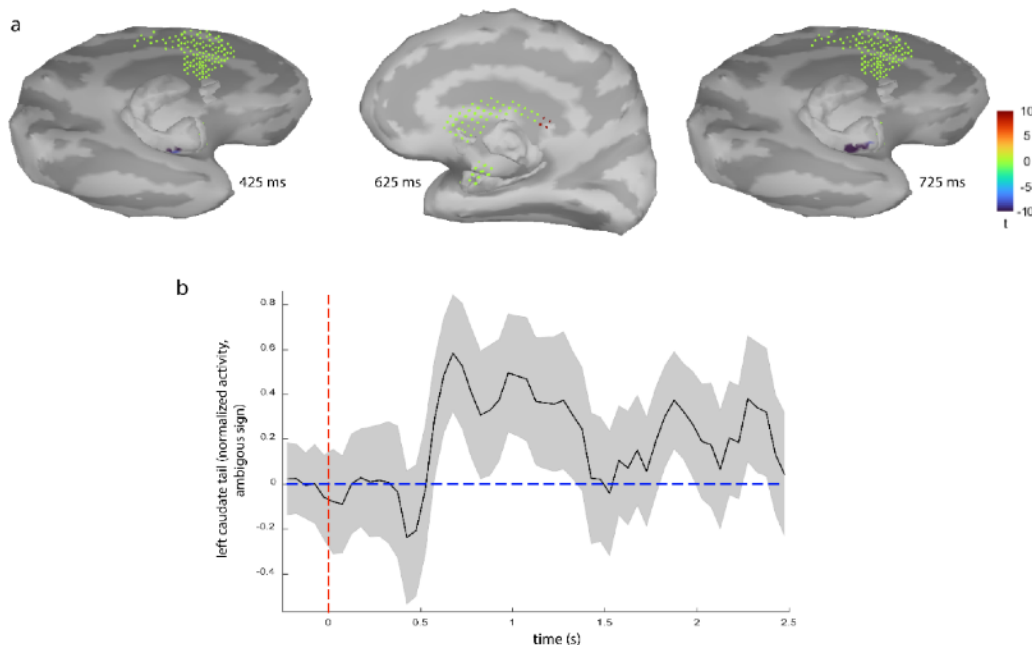
Figure 3: EEG sources of the value ERP. a) The t-stat maps for the group average of the sources of value ERP, against their baseline activities (two-sided t-test; p-value < 0.001, duration > 150ms, cluster > 3 dipoles). Note that subcortical regions with unknown dipole directions (caudate & amygdala) are shown as masses of their grid points along with the opposite-side hemispheres. This map shows activation of the left caudate tail and deactivation of anterior hippocampus (note the ambiguity in the sign of dipole activities in caudate). **b)** The group-average time course of the activity in the caudate tail (normalized magnitude) along with standard error margins indicating group diversity around the mean.

1 SNR: Signal to Noise Ratio

Functional connectivity between CDt and FFA

How does CDt receive the specific information of each face? To address this question, we performed a functional connectivity analysis (based on GLM1 residuals) to find the probable cortical or subcortical sources of this face-specific information, by selecting the detected left CDt in GLM1 (Fig. 3b) as the correlation seed (see methods for details). Multiple regions in the temporal and frontal cortex were found to be functionally connected to the left CDt (Fig. 4a: p-value < 0.001, cluster-corrected, suppl. Table 2). Interestingly, face processing areas in the inferior temporal cortex were among the functionally connected regions. In particular, the anterior part of the left fusiform gyrus (FFA) found at the group-level by the face localizer scans performed for each subject (faces vs. scrambled faces), overlapped with the functionally connected areas to CDt (Fig. 4a; p-value<0.001, cluster-corrected, see methods & suppl. Fig. 7). Despite this overlap, FFA itself does not show a significant differentiation between remembered good and bad faces (GLM2, Fig.4b), suggesting that while face-specific information can be supplied to CDt by FFA, it is the CDt which does the moral value-based face discrimination.
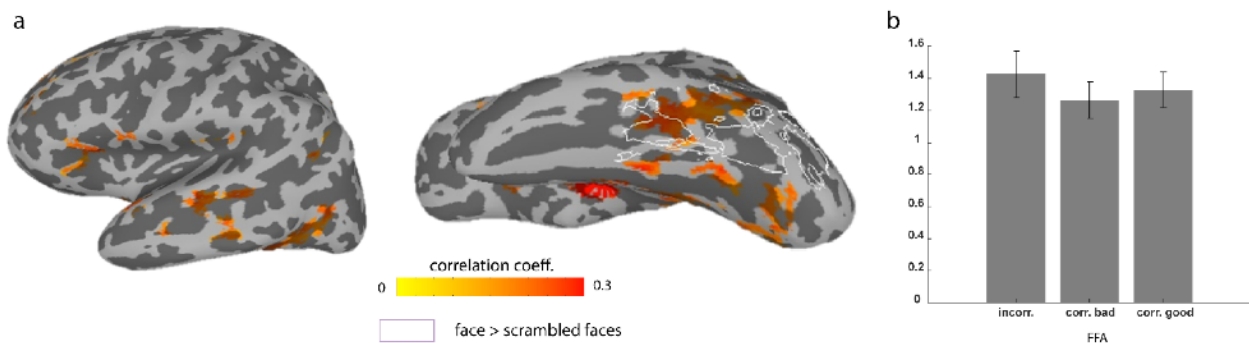
Figure 4: Functionally connected regions to the left CDt overlap with FFA. a) functional connectivity map for the left caudate tail on the left hemisphere: the group-average of the correlation coefficients (p-value < 0.001, cluster-corrected). Contours of cortical face-selective areas (using a separate face localizer scan for each subject) are marked. **b)** The beta values for correct good, correct bad and incorrect answers in the right fusiform gyrus (averaged over cluster and subjects), compared in bar plots. See

Suppl. Fig. 7 for more details about the additional localizer task used to localize the FFA region for each subject separately.

Discussion:

Our social interactions are highly affected by our judgements about a person's integrity. In many situations one or two encounters with events that show ethical or unethical behavior of a person, is enough for us to form lasting positive or negative memories of that individual. Yet, the neural mechanism of such a robust phenomenon was not previously addressed. Here, we investigated the neural encoding of long-term memory of moral values associated with human faces using simultaneous EEG-fMRI data acquisition to reveal both spatial and temporal dynamics of brain activations. First, our behavioral results confirmed that only a few exposures to morally charged stories of people are adequate for forming long-term memories a day later (Fig. 1). There was significant differentiation of the memorized good vs bad face ERPs over center-frontal electrodes lasting during the face presentation (value ERP). EEG-informed fMRI analysis using the power of EEG signal over the center-frontal electrodes revealed a significant activity centered on the left CDt (Fig 2). Conversely, fMRI-informed EEG source reconstruction localized sources of "value ERP" in CDt with an onset time of about 550ms (Fig. 3). Notably, EEG-informed fMRI analysis and fMRI-informed EEG source localization as well as traditional whole brain fMRI analysis did not show any significant differentiation of remembered good and bad faces in any cortical areas including the face processing regions (Table 1). Nevertheless, functional connectivity analysis revealed a connection between CDt and anterior FFA which presumably can be the source of face specific information to this part of the caudate (Fig. 4). The subjects had a higher performance in detecting good faces compared with bad ones. While a bias toward the positive value in some subjects could probably be a partial reason for this observation, this does not alter our findings since they rely only on the correctly chosen values for good and bad faces. Furthermore, the subjects' response statistics for each face also confirmed that there was no value bias in most of the faces and exclusion of the few biased faces did not change our results (suppl. Fig. 5).

Caudate and in particular its tail region was previously shown to be a key node for encoding long-term value memories of objects in general, in non-human primates (Kim & Hikosaka, 2013; Yamamoto et al., 2013) and in humans (Farmani Sepideh et al., n.d.). Single-unit recordings from monkey caudate tail has showed higher activation to good compared to bad objects (Yamamoto et al., 2013). Our results extend these previous findings by implicating CDt in differentiation of faces based on good and bad moral values (Fig. 2,3 & suppl. Fig.4). Electrophysiological recordings and fMRI data from monkeys have shown several cortical regions to be involved with object value memory including areas in temporal and prefrontal cortices (Ghazizadeh, Griggs, et al., 2018; Ghazizadeh, Hong, et al., 2018). While we did not find significant cortical

representation of long-term value memory in those areas, we found some of them (vlPFC & STS, Fig. 4a, suppl. Table 2) functionally connected to the left caudate tail. We also note that the previously found cortical activations were observed for over-trained objects (>10 days reward learning) while in our tasks the value of each faces were only encountered 2 times.

Some recent EEG and fMRI studies have assigned social/moral values to faces independent of their physical attributes. While (Cloutier & Gyurovski, 2014; Luo et al., 2019) have studied the fMRI BOLD changes during the training or value assignment phase and shown activities in the ventromedial prefrontal cortex (vmPFC), middle occipital gyrus (mOG) and mOFC, (Singer et al., 2004; Todorov et al., 2007) reported differential BOLD responses for valued faces compared with neutral faces in various regions including insula, fusiform gyrus and STS. Some EEG studies have also paired faces and morally-charged stories, and show an enhancement of the late positive potential (LPP, 400~600ms post-stimulus) for the valued faces compared to neutral ones, as well as a weaker effect for the N170 and early posterior negativity (EPN, 250~350ms) components (Baum & Abdel Rahman, 2021b, 2021a; Schindler et al., 2021; Suess et al., 2015). None of these studies however, address neural mechanism for long-term storage of facial moral values. This is particularly important in the light of literature that shows a clear segregation of short- and long-term value memories for objects (Kim paper). Notably, recent studies on long-term memory of non-face object values in humans (Farmani Sepideh et al., n.d.; Kang et al., 2021) have shown value discrimination in ventral striatum and part of the caudate. The absence of activation in ventral striatum in our study could to be due to the difference in stimuli type (the faces vs non-face objects) or to the form of value (money reward for objects vs moral values for faces). Further investigations are needed to check whether long-term value memory are encoded in different parts of the striatum based on the stimulus and reward types.

While the onset of face value ERP in center-frontal scalp EEGs and in caudate tail activation detected by EEG source reconstruction (~550ms, Fig. 2,3) seems compatible with the onset of value-representing ERP signal in (R. Frömer et al., 2023) and the onset of value-related LPP components in (Baum & Abdel Rahman, 2021b; Schindler et al., 2021), it seems to be later than the onset of value signal in the electrophysiological recordings from monkeys in the caudate tail (~ 150ms) (Yamamoto et al., 2013). This could be due to difference in stimuli used (fractals vs face) and the type and duration of value training in the two experiments if not the species differences. We also note that the value ERP also showed a smaller yet significant (p<0.001, but not corrected) value differentiation around 180ms in FC1 electrode.

We also note that while both EEG-informed fMRI analysis and fMRI-informed EEG source localization found CDt as the main substrate for encoding remembered good and bad faces, there is a mismatch in the exact location of CDt in the two methods. This is mainly due to the limitations in EEG source localization and its subcortical head model for the caudate which does

not include the CDt part immediately adjacent to the hippocampus. Nevertheless, the fMRI-informed EEG source localization still manages to find the most posterior part of CD in its model which is part of CDt, as the source of face value differentiation. Notably, this analysis finds no value encoding neither in other subcortical areas (but some deactivation in the hippocampus) nor in the cortical areas despite their much shorter distance to the EEG electrodes consistent with lack of such activation in EEG-informed fMRI and the traditional whole brain fMRI analyses.

Interestingly, we did not find any significant activation of face processing areas to moral values of faces seen a day before. The fusiform gyrus is previously shown to encode facial expressions to some degrees (Engell & Haxby, 2007; Winston et al., 2004). (Singer et al., 2004) assigned social value to faces through a prisoner's dilemma game and also showed an increased activity in the fusiform gyrus for good faces compared to neutral ones, albeit shortly after the value assignment. This modulation is more likely to reflect an elevated salience for good faces rather than a coding of face value. Indeed, recent ERP analyses showed enhanced LPP component for *both* good and bad faces, indicating an effect of saliency for face values rather than a value coding (Baum & Abdel Rahman, 2021b; Schindler et al., 2021). Together, our findings and these results suggest that while value maybe encoded in caudate, salience of faces may engage a wider circuitry including the face processing areas.

In summary, our results showed robust coding of face moral values in the CDt. Functional connectivity analysis showed this part of caudate to be connected to FFA, which can provide the face specific information. Similar to other parts of striatum, CDt is a major target for dopaminergic (DA) neurons, in particular its posterior subpopulation that is known to differentiate objects based on their values (Kim et al., 2015). The exact mechanism by which this posterior basal ganglion circuitry work to differentiate faces based on their moral values in interaction with other functionally connected cortical and subcortical areas remains to be addressed.

Materials & Methods:

participants:

Twenty-one subjects (16 males, 5 females), aged between 21 and 40 years (mean=26.6 years, s.d.±5.3) participated in the experiment. They were all healthy and right-handed and had normal or corrected-to-normal vision. Written informed consent was obtained in accordance with the School of Cognitive Sciences Ethics Committee at the Institute for research in fundamental sciences (IPM) in Tehran, ethics code: 99/60/1/6117

Stimuli:

We used 24 arbitrary artificial faces (13 males, 11 females) created by a deep neural network model, the StyleGAN2 (https://thispersondoesnotexist.com) (Karras et al., 2019). Each face was randomly assigned to a brief unique biography with either a positive (good face) or a negative (bad face) moral value (see suppl. Table 1 for the list of all the stories). Assignment of stories to faces were random and the positive or negative values were swapped across subjects.

Training session:

The night before the experiment, the subjects watched a short (5 minutes) video (video 1 or 2) which introduced 24 faces with a brief biographical history about each face narrated in Persian (the subjects' native language). Each history included a positive or a negative ethical value for the corresponding face on the screen (good and bad faces, Fig. 1a). The subjects passively watched the video two times, once at 7pm and the second time at 9pm the night before the experiment. Each face was portrayed for 10s and the narration of history started with the emergence of the face. For a complete list of the histories translated to English, see suppl. Table 1. The memory session started about 10:00 AM the next day thus testing the long-term memory across a 13~14 hours timespan.

Memory session:

During the memory session, the subjects were first equipped with an MRI-compatible EEG cap and then laid on the MRI bed. They were also given two response handles to each hand to indicate their responses by pressing the buttons using their index fingers. The faces introduced in the training video were shown in a random order for each subject and he/she was instructed to indicate his/her judgement about the moral value of the presented face. Each trial started with a fixation cross for a random period between 1 to 3.5 s. Then a face was portrayed for 2.5s and the subject passively watched it. After that, the face disappeared and two letters of 'G' and 'B' (referring to "good" and "bad" respectively) were shown in the left and right sides of the midpoint of the screen for 2s, during which the subject had to indicate his/her response. The sides of the letters 'G' and 'B' were randomly flipped in each trial. The subjects were instructed to answer for all faces even if they did not remember the story of the face or its exact value. Each face was shown only once during this test and the subjects performed only one run of this experiment. Subjects made a choice for almost all presented faces (99%).

fMRI data acquisition:

We acquired our fMRI data using a 3T 'Siemens' scanner in the "National Brain Mapping Laboratory, NBML" in Tehran. Specifically, we collected functional Echo-Planar-Imaging (EPI) data using a 64-channel head coil with an anterior–posterior fold over direction (repetition time: 2.5 s; echo time: 30ms; number of slices: 42; number of voxels: 70×70; in-plane resolution: 3.543×3.543mm; slice thickness: 3.5mm; flip angle: 80º). Slices were collected in an interleaved order. Anatomical images were acquired using a MPRAGE T1-weighted sequence that yielded images with a 1×1×1mm resolution (176 slices; number of voxels: 256×256; repetition time: 2000ms; echo time: 3.47ms) as well as a T2 image with a 0.9×0.475×0.475mm resolution (192 slices; number of voxels: 512×512; repetition time: 3200ms; echo time: 408ms). We also acquired field-map gradient images using a multi-shot gradient echo sequence which was subsequently used to correct for distortions in the EPI data due to B0 inhomogeneities (echo times: TE1=4.92ms, TE2= 7.38ms; isotropic resolution: 3.75mm; matrix: 64×64×38; repetition time: 476ms; flip angle: 60º).

fMRI pre-processing:

We discarded the first three volumes from each fMRI run (due to magnetization artifact). We performed the pre-processing steps using FSL. These steps include motion-correction, field-map correction, slice-time correction, high-pass filtering (>100 s) and spatial smoothing to 5mm. The EPI images of each subject were first registered to his/her structural image using the BBR algorithm. Registration of structural images to the MNI brain was performed using the nonlinear method with a 10 mm warp resolution.

Power analysis:

We conducted a power analysis based on GLM1 of the first 7 subjects and using the GPower software with parameters: two-sided t-test, alfa=0.05 and power=0.85 which resulted in a minimum number of 14 subjects.

fMRI GLM Analysis:

We used FSL to run the GLM analyses. We used the pre-whitening option and the 3-column or 1-column format for regressor specification. In particular, each of the four main regressors was built as boxcars over the corresponding time intervals. For example, the "correct good" regressor was equal to 1 during all the 2.5s exposures to the good faces that were correctly remembered as good. The two confounding regressors for hand responses were built as stick functions (100ms wide and amplitude equal to 1) at 100ms before the response time (accounting for motor delay), for subjects with accurately saved response times (7 subjects). For the rest of the subjects, the reaction times were not saved. For these subjects we used the average reaction time of those seven subjects and placed 600ms wide boxcars (with 1/6 height) centered at the average reaction time. The 600ms was chosen as twice the standard deviation (std) of the saved reaction times (300ms). Note that the subjects had to answer within time slots of 2s which was smaller than one fMRI repetition time. Then, we convolved all these regressors with the canonical hemodynamic response function (HRF) to be used in the GLM analysis. We also used a ventricle mask (in the native space of each subject) to extract the non-neuronal time-series of the BOLD signal inside the ventricles and used the resulting signal as a confound regressor. In our second GLM analysis (EEG-informed fMRI) we also calculated the average of electric potentials on the electrodes C1 and FC1, and then its instantaneous power. We used the EEG power in intervals 0.5 to 1.5s post-stimulus and set the signal outside these intervals equal to zero. Convolution of the resulting signal with HRF, gave us our EEG-driven regressor. In particular, the GLM model was:

$$y = \beta X + r \tag{1}$$

where, Y is the time series of the preprocessed and demeaned BOLD response of a single voxel for T time samples, X is a $n \times T$ design matrix with rows representing n regressors (here, n=7: 4 main +3 confounding for our ordinary GLM, and n=8 for our EEG-informed fMRI analysis), β is a $1 \times n$ vector containing the regression weights for each regressor for this particular voxel and r is the $1 \times T$ residual of this regression. The resulting regression coefficients (beta maps) were finally normalized by the temporal mean of subject's EPI images.

fMRI group-level Analysis and statistical tests:

We performed the group-level and statistical analyses using functions from AFNI 20.2.05. For group-level analysis, we used the function "3dttest++". The group-level activation maps were then masked by the grey matter mask associated with the standard MNI brain with a resolution of 2mm. By applying 3dFWHMx on these group-level residuals, we estimated the parameters for the non-Gaussian spatial autocorrelation function of the fMRI noise. 3dClustSim was used to calculate the cluster thresholds for

various *p*-values such that the probability of a false positive cluster among the *p*-thresholded clusters was less than α=0.05. Statistical corrections for subcortical sources were calculated using a subcortical gray matter mask, due to their intrinsic smaller sizes. The inflated surfaces for visualization of fMRI results are presented using SUMA 20.2.05.

EEG data acquisition:

EEG data was acquired at a 5-kHz sampling rate at the same time as the fMRI data collection, using an MR-compatible EEG amplifier system (BrainAmps MR-Plus, Brain Products, Germany) and the Brain Vision Recorder software (BVR; Version 1.10, Brain Products, Germany). Data were filtered online with a hardware band-pass filter of 0.1 to 250Hz. The EEG cap included 65 Ag/AgCl scalp electrodes which were localized according to the international 10–20 system. The AFz and FCz electrodes were chosen as the ground and reference electrodes, respectively. All electrodes had in-line 10k surface-mount resistors to ensure subject safety. All leads were bundled together and twisted for their entire length to minimize inductive pick-up and ensure participants' safety. Input impedances were kept below 20k (including the 10k surface-mount resistors). EEG data acquisition was synchronized with the fMRI data (Syncbox, Brain Products, Germany) and triggers from the MR-scanner were collected separately to remove MR gradient artefacts offline.

Recording the EEG electrode coordinates:

We captured the EEG electrode coordinates registered to the subjects T1-image using 'Localite' TMS navigator just after when subject came out of the MRI scanner.

EEG pre-processing:

We performed EEG-preprocessing using the EEGLAB toolbox in MATLAB. MRI gradient noise was removed via the FMRIB's "FASTER" plug-in for EEGLAB. After down-sampling the resulting signal to 1000 Hz, the signal was bandpass filtered between 0.5Hz to 40 Hz. Then, we detected the QRS events from the ECG signal and removed the ballistocardiogram (BCG) artifact using the FMRIB plug-in for EEGLAB. Subsequently, we detected the high noise sections or channels in the data by visual inspection and put them equal to zero and then, used ICA decomposition in order to remove the eye-motion artifacts, the residual of BCG and other non-brain sources. Finally, we interpolated high noise sections or channels (if any) and re-referenced the EEG signals to the common average so that we could recover the signal at FCz. For Event-Related-Potentials (ERPs) we also subtracted the average of the signal over the baseline period (250 ms before stimulus). Four subjects (out of the whole 21 subjects) were excluded from further EEG analyses due to their severe residual MRI gradient noise. One subject was also excluded due to not showing meaningful visual ERPs. The group-average of the ERPs in electrode O1 (suppl. Fig. 1) showed a strong visual ERP with clear P200 and P300 components, ensuring a suitable removal of MRI gradient noise from the EEG signals.

EEG source reconstruction:

We used "Brainstorm" toolbox in MATLAB for EEG source localization. We conducted the MRI segmentations based on both structural T1 and T2 MRI images of each subject using "FreeSurfer" program and generated the boundary element model (BEM) based on these segmentations. We used the simple BEM model to register the scalp potentials to the standard MNI scalp (described further in detail).

To localize the sources of the value ERP, we needed to consider the subcortical areas in our forward model as well, since the fMRI analysis had shown the caudate nucleus to encode the value memory of the faces. In order to get more accurate localization results, we used a "mixed" head model offered by Brainstorm. Mixed models do not place dipoles in the white matter and assume dipoles on the cortex surface to be perpendicular to it, but let subcortical dipoles to have unknown directions. We avoided using a finite-element model (FEM) due to the very large number of unknown variables considered for the inverse problem in this model, which can severely degrade the results of the ill-posed source reconstruction problem. In the "mixed" forward model offered by Brainstorm, a constrained BEM model for the cortex (one dipole at each grid, perpendicular to the cortex) is combined with an unconstrained model for subcortical grey matter regions (three orthogonal dipoles at each subcortical grid). Some subcortical regions (e.g. hippocampus) also have known dipole directions which are considered in this mixed model.

The time-series of neuronal activities were reconstructed using weighted minimum norm estimation (wMNE), and the "automatic shrinkage" method for noise covariance matrix regularization. We estimated the noise covariance matrix based on the rest periods of the experiment.

In order to boost the signal-to-noise ratio of the estimated source time series and prepare them for the group-level analysis, we divided the ERP time course (+ its baseline) into 55 time-bins of 50 ms and conducted the temporal mean of each source over each time bin. Then we z-scored all source time series by subtracting the baseline mean and dividing by baseline std (for the subcortical sources, division by the norm of the 3D activities). In order to figure out a single activation magnitude for subcortical sources, we computed the principle component of the associated three time-vectors at each grid point, which revealed the main dipole directions but for a sign ambiguity.

Cluster-correction for EEG sources:

In order to address the multiple-comparison problem for the detected EEG sources, we conducted cluster thresholds similar to that used for fMRI results. Specifically, we mapped the baseline activities of the EEG sources to the MRI voxels including them and then used the 3dFWHMx function to estimate the spatial autocorrelation parameters of the noise data. Then we used 3dClustSim function to calculate the cluster thresholds so that the false positive rate was kept below 5%. We did this paradigm separately for the cortical and subcortical sources. We restricted these calculations to a caudate mask for the detected activity in the CDt. The equivalent volume for the subcortical threshold was simply calculated using Brainstorm GUI. For the cortical surface sources, the equivalent area of the threshold was conducted and the number of vertices building this area was to be estimated again using Brainstorm GUI. Due to non-homogeneous parcellation of the cortex surface we selected the maximum vertex cluster size building up the same thresholding area.

Calculation of the group-average ERPs:

Given the variations in the subjects' head size and shape as well as some possible and unavoidable variations in EEG cap placement on individual subjects' heads, it would be better to map the EEG topographies and ERPs on a common standard head in order to compute a more precise group average of them. This implies the access to the exact position of the EEG electrodes for each subject. We had recorded these exact electrode coordinates for each subject, in order to conduct a more precise lead-field matrix. Therefore, we used this information to map single-subject ERPs to a standard scalp. Specifically, we first mapped ERPs to brain sources in the native space and then mapped them to the standard MNI brain and finally mapped these activities back to the surface potentials on the MNI scalp.

Since we used a linear inverse solution, the overall end-to-end mapping of ERPs to the standard head is a linear mapping and hence, it is completely reversible. The resultant potentials were then normalized by the overall norm of channels' baselines activities for each subject, before averaging. Also note that we used this mapping and normalization only to conduct the average of value ERPs across subjects, while raw EEG signals and raw ERPs were used in the subject-level "EEG-informed fMRI GLM" and "EEG source reconstruction" analyses respectively.

Evaluation of the proposed method for mapping subject ERPs on a standard scalp:

As mentioned above, the proposed method is a linear reversible interpolation method for mapping the subject ERPs over another (standard) scalp. The main concern is to preserve the overall scalp topography and thus, some probable errors in the source reconstruction are not important unless they affect the scalp topography. Such a localization error does not result in any changes to the scalp topography in the native space (note the under-determinacy of the localization problem) and hence it belongs to the null-space of the native lead-field matrix. The intuition is that these errors would also result in negligible deflections in scalp topographies on the standard scalp, given the high correlations between the rows of the native lead-field matrix and the lead-field matrix for a standard brain. In order to quantify our intuition, we ran simulations by assigning neuronal dipoles with random magnitudes according to a normal distribution. Then we reconstructed the sources based on MNE method, mapped them to the MNI brain and consequently over the MNI scalp and then conducted the correlations between the EEG signals (topographies) in the native and standard spaces. We repeated these simulations 1000 times. The resulted histogram of these correlations indicates that this correlation was greater than 0.5 and of course it was often about 0.9 (suppl. Fig 8a).

For the real EEG, our method has an extra added value of denoising the scalp topographies. Using a noise covariance matrix weighting method inside the wMNE solution, the effects of noisy channels are highly alleviated. We also note that in tackling with real data, the EEG noise covariance matrices (or simply the set of noisy channels) differ among subjects, and hence this is an added value of our mapping method. In order to evaluate the performance of this method on real data, we took a visual ERP of the subject, partitioned it to intervals of 50ms and conducted the time-average over each interval. We then conducted the correlations between the denoised EEG topography in the native space and that mapped over the standard scalp. Similar to the synthetic data, the correlations between the denoised EEG in the native and standard spaces were high and often about 0.9 (suppl. Fig 8b).

FFA Localizer task:

In a separate experimental run, the subjects passively watched blocks of four image types: general objects, the scrambled version of the same objects, human faces (other than those introduced in the training video) and the scrambled versions of the same faces, being repeated in three periods (suppl. Fig. 7a). Each block lasted for 12.5s and contained images of 10 different random samples of that category chosen from a source of 100 images. Each sample image was portrayed for 1s, followed by a 250ms inter-stimulus-interval. We analysed the BOLD signals using a GLM with four regressors, one for each category. The fusiform face area was localized for each subject by thresholding the "face vs scrambled face" contrast with p-value < 10^{-5}. The group average of this contrast showed activities in the FFA and OFA (suppl. Fig. 7b, p-value < 10^{-4}, cluster-corrected).

The functional connectivity analysis:

We used the residuals from our first GLM (the EEG-informed fMRI) analysis as an estimate of the non-modulated brain activity. We used the detected CDt cluster in GLM1 (Fig. 2c) as the correlation seed. We

Code Accessibility:

The AFNI codes used for fMRI analyses as well as MATLAB codes used for EEG analyses are available at: https://github.com/AliAtaei1/face-value

Acknowledgements:

Table 1: activated clusters. Sizes, coordinates and cluster-correction thresholds for the activated clusters. CM: Center of Mass

	cluster size (vox-...)	CM x	CM y	CM z	peak x	peak y	peak z	cluster threshold for p=0.01 (vox-...)
corr. good -corr. Bad (GLM2):								66
L caudate tail	86	29.3	38.3	-1.2	32	38	-4	
EEG-driven regressor (GLM1):								54
L caudate tail	55	29.8	27.5	-6.1	28	26	-4	
EEG-driven regressor (GLM3):								56
L caudate tail	129	29.5	28.7	-8.8	22	26	-10	
correct - incorrect:								74
L angular gyrus	296	47.7	59.5	35.1	46	56	34	
L superior frontal gyrus	151	18.7	-27.8	52.2	18	-28	58	
L medial temporal gyrus	75	64	25.9	-7.3	62	26	-8	
right hand - left hand								80
R motor cortex	1393	-37.8	25	56.1	-52	16	56	
L cerebellum	655	17.5	52.7	-21	26	50	-22	

L lingual cortex	583	19.2	65	-8.9	16	74	-10
R insula	447	-45.6	16.4	13.9	-48	14	16
R cerebellum	328	-20.6	48.9	-22.5	-20	50	-22
L motor cortex	318	36.9	23.3	57.5	36	22	44
L cerebellum2	261	16.3	60.3	-47	16	58	-46
R cingulate gyrus	254	-7.6	10.4	50	-10	26	48
R posterior cingulate	242	-12.7	59.8	3.6	-4	74	8
R Cuneus	215	-13.1	93.7	4	-16	94	0
L Cuneus	200	19	86.3	23.9	12	86	28
L Cuneus2	164	8	76.8	9.5	10	70	6
R thalamus	149	-10.9	18.4	9.7	-10	20	12
R superior temporal gyrus	117	-55.5	16.1	7.6	-50	18	10
L Cuneus3	112	3	87.4	10.5	0	90	18
R inferior frontal gyrus	84	-59.7	-7.3	17.9	-60	-12	22

References:

Ataei, A., Amini, A., & Ghazizadeh, A. (2022). Gustatory Cortex Is Involved in Evidence Accumulation during Food Choice. *Eneuro*, *9*(3), ENEURO.0006-22.2022. https://doi.org/10.1523/ENEURO.0006-22.2022

Baum, J., & Abdel Rahman, R. (2021a). Emotional news affects social judgments independent of perceived media credibility. *Social Cognitive and Affective Neuroscience*, *16*(3), 280–291. https://doi.org/10.1093/scan/nsaa164

Baum, J., & Abdel Rahman, R. (2021b). Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation. *NeuroImage*, *244*, 118572. https://doi.org/10.1016/j.neuroimage.2021.118572

Berridge, K. C., & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology*, *199*(3), 457–480. https://doi.org/10.1007/s00213-008-1099-6

Cloutier, J., & Gyurovski, I. (2014). Ventral medial prefrontal cortex and person evaluation: Forming impressions of others varying in financial and moral status. *NeuroImage*, *100*, 535–543. https://doi.org/10.1016/j.neuroimage.2014.06.024

Cloutier, J., Heatherton, T. F., Whalen, P. J., & Kelley, W. M. (2008). Are Attractive People Rewarding? Sex Differences in the Neural Substrates of Facial Attractiveness. *Journal of Cognitive Neuroscience*, *20*(6), 941–951. https://doi.org/10.1162/jocn.2008.20062

Cornwell, J. F. M., & Higgins, E. T. (2019). Sense of Personal Control Intensifies Moral Judgments of Others' Actions. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02261

Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science*, *1*(1), 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, *45*(14), 3234–3241. https://doi.org/10.1016/j.neuropsychologia.2007.06.022

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*(9), 1508–1519. https://doi.org/10.1162/jocn.2007.19.9.1508

Farmani Sepideh, Sharifi Kiomars, & Ghazizadeh Ali. (n.d.). Cortical and subcortical substrates of minutes and days-long object value memory in humans. *BioRxiv*. https://doi.org/https://doi.org/10.1101/2023.03.17.533079

Ghazizadeh, A., Griggs, W., Leopold, D. A., & Hikosaka, O. (2018). Temporal–prefrontal cortical network for discrimination of valuable objects in long-term memory. *Proceedings of the National Academy of Sciences*, *115*(9). https://doi.org/10.1073/pnas.1707695115

Ghazizadeh, A., & Hikosaka, O. (2021). Common coding of expected value and value uncertainty memories in the prefrontal cortex and basal ganglia output. *Science Advances*, *7*(20), eabe0693. https://doi.org/10.1126/sciadv.abe0693

Ghazizadeh, A., Hong, S., & Hikosaka, O. (2018). Prefrontal Cortex Represents Long-Term Memory of Object Values for Months. *Current Biology*, *28*(14), 2206-2217.e5. https://doi.org/https://doi.org/10.1016/j.cub.2018.05.017

Greene, J. D. (2009). The Cognitive Neuroscience of Moral Judgment. In *The Cognitive Neurosciences*. The MIT Press. https://doi.org/10.7551/mitpress/8029.003.0086

Jiang, Q., Zhuo, L., Wang, Q., & Lin, W. (2022). The Neural Basis of Moral Judgement for Self and for Others: Evidence From Event-Related Potentials. *Frontiers in Human Neuroscience*, *16*. https://doi.org/10.3389/fnhum.2022.919499

Kang, J., Kim, H., Hwang, S. H., Han, M., Lee, S.-H., & Kim, H. F. (2021). Primate ventral striatum maintains neural representations of the value of previously rewarded objects for habitual seeking. *Nature Communications*, *12*(1), 2100. https://doi.org/10.1038/s41467-021-22335-5

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019). *Analyzing and Improving the Image Quality of StyleGAN*.

Kim, H. F., Ghazizadeh, A., & Hikosaka, O. (2014). Separate groups of dopamine neurons innervate caudate head and tail encoding flexible and stable value memories. *Frontiers in Neuroanatomy*, *8*. https://doi.org/10.3389/fnana.2014.00120

Kim, H. F., Ghazizadeh, A., & Hikosaka, O. (2015). Dopamine Neurons Encoding Long-Term Memory of Object Value for Habitual Behavior. *Cell*, *163*(5). https://doi.org/10.1016/j.cell.2015.10.063

Kim, H. F., & Hikosaka, O. (2013). Distinct Basal Ganglia Circuits Controlling Behaviors Guided by Flexible and Stable Values. *Neuron*, *79*(5), 1001–1010. https://doi.org/10.1016/j.neuron.2013.06.044

Kranz, F., & Ishai, A. (2006). Face Perception Is Modulated by Sexual Preference. *Current Biology*, *16*(1), 63–68. https://doi.org/10.1016/j.cub.2005.10.070

Luo, Q., Yu, M., Li, Y., & Mo, L. (2019). The neural correlates of integrated aesthetics between moral and facial beauty. *Scientific Reports*, *9*(1), 1980. https://doi.org/10.1038/s41598-019-38553-3

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *8*(3), 285–299. https://doi.org/10.1093/scan/nsr090

Michel, C. M., & Brunet, D. (2019). EEG Source Imaging: A Practical Review of the Analysis Steps. *Frontiers in Neurology*, *10*. https://doi.org/10.3389/fneur.2019.00325

Michel, C. M., & He, B. (2019). *EEG source localization* (pp. 85–101). https://doi.org/10.1016/B978-0-444-64032-1.00006-0

Morrison, S. E., & Salzman, C. D. (2010). Re-valuing the amygdala. *Current Opinion in Neurobiology*, *20*(2), 221–230. https://doi.org/10.1016/j.conb.2010.02.007

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, *41*(2), 147–155. https://doi.org/10.1016/S0028-3932(02)00145-8

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226. https://doi.org/10.1038/nature04676

R. Frömer, M.R. Nassar, B.V. Ehinger, & A. Shenhav. (2023). *Common neural choice signals emerge artifactually amidst multiple distinct value signals*. bioRxiv.

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*(4), 389–397. https://doi.org/10.1038/nn2066

Said, C. P., Haxby, J. V., & Todorov, A. (2011). Brain systems for assessing the affective value of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1660–1670. https://doi.org/10.1098/rstb.2010.0351

Schindler, S., Bruchmann, M., Krasowski, C., Moeck, R., & Straube, T. (2021). Charged With a Crime: The Neuronal Signature of Processing Negatively Evaluated Faces Under Different Attentional Conditions. *Psychological Science*, *32*(8), 1311–1324. https://doi.org/10.1177/0956797621996667

Schultz, W. (2007). Multiple Dopamine Functions at Different Time Courses. *Annual Review of Neuroscience*, *30*(1), 259–288. https://doi.org/10.1146/annurev.neuro.28.061604.135722

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*(4), 653–662. https://doi.org/10.1016/S0896-6273(04)00014-5

Suess, F., Rabovsky, M., & Abdel Rahman, R. (2015). Perceiving emotions in neutral faces: expression processing is biased by affective person knowledge. *Social Cognitive and Affective Neuroscience*, *10*(4), 531–536. https://doi.org/10.1093/scan/nsu088

Thickbroom, G. W., Byrnes, M. L., & Mastaglia, F. L. (2003). Dual representation of the hand in the cerebellum: activation with voluntary and passive finger movement. *NeuroImage*, *18*(3), 670–674. https://doi.org/10.1016/S1053-8119(02)00055-1

Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*(1), 163–173. https://doi.org/10.1016/j.neuropsychologia.2006.04.018

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*(10), 2766–2781. https://doi.org/10.1162/jocn.2011.21616

Wassum, K. M., & Izquierdo, A. (2015). The basolateral amygdala in reward learning and addiction. *Neuroscience & Biobehavioral Reviews*, *57*, 271–283. https://doi.org/10.1016/j.neubiorev.2015.08.017

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-Adaptation Reveals Dissociable Neural Representations of Identity and Expression in Face Perception. *Journal of Neurophysiology*, *92*(3), 1830–1839. https://doi.org/10.1152/jn.00155.2004

Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, *45*(1), 195–206. https://doi.org/10.1016/j.neuropsychologia.2006.05.009

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*(3), 277–283. https://doi.org/10.1038/nn816

Yamamoto, S., Kim, H. F., & Hikosaka, O. (2013). Reward value-contingent changes of visual responses in the primate caudate tail associated with a visuomotor skill. *Journal of Neuroscience*, *33*(27), 11227–11238. https://doi.org/10.1523/JNEUROSCI.0318-13.2013

Yasuda, M., Yamamoto, S., & Hikosaka, O. (2012). Robust representation of stable object values in the oculomotor Basal Ganglia. *Journal of Neuroscience*, *32*(47), 16917–16932. https://doi.org/10.1523/JNEUROSCI.3438-12.2012