Neural interactions in the human frontal cortex dissociate reward and punishment learning

Etienne Combrisson[1], Ruggero Basanisi[1], Maëlle C. M. Gueguen[2], Sylvain Rheims[3], Philippe Kahane[4], Julien Bastin[2], Andrea Brovelli[1]

[1] Institut de Neurosciences de la Timone, Aix Marseille Université, UMR 7289 CNRS, 13005, Marseille, France
[2] Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN, Grenoble, France
[3] Department of Functional Neurology and Epileptology, Hospices Civils de Lyon and University of Lyon, Lyon, France
[4] Univ. Grenoble Alpes, Inserm, U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, GIN, Grenoble, France

Abstract

How human prefrontal and insular regions interact while maximizing rewards and minimizing punishments is unknown. Capitalizing on human intracranial recordings, we demonstrate that the functional specificity toward reward or punishment learning is better disentangled by interactions compared to local representations. Prefrontal and insular cortices display non-selective neural populations to reward and punishment. The non-selective responses, however, give rise to context-specific interareal interactions. We identify a reward subsystem with redundant interactions between the orbitofrontal and ventromedial prefrontal cortices, with a driving role of the latter. In addition, we find a punishment subsystem with redundant interactions between the insular and dorsolateral cortices, with a driving role of the insula. Finally, switching between reward and punishment learning is mediated by synergistic interactions between the two subsystems. These results provide a unifying explanation of distributed cortical representations and interactions supporting reward and punishment learning.

Introduction

Reward and punishment learning are two key facets of human and animal behavior, because they grant successful adaptation to changes in the environment and avoidance of potential harm. These learning abilities are formalized by the law of effect (Thorndike, 1898; Bouton, 2007) and they pertain the goal-directed system, which supports the acquisition of action-outcome contingencies and the selection of actions according to expected outcomes, as well as current goal and motivational state (Dickinson and Balleine, 1994; Balleine and Dickinson, 1998; Balleine and O'doherty, 2010; Dolan and Dayan, 2013; Balleine, 2019).

At the neural level, a first hypothesis suggests that these abilities are supported by distinct brain circuits (Pessiglione and Delgado, 2015; Palminteri and Pessiglione, 2017). Indeed, an anatomical dissociation between neural correlates of reward and punishment prediction error (PE) signals has been observed. PE signals are formalized by associative models (Rescorla et al., 1972)and reinforcement learning theory (Sutton and Barto, 2018) as the difference between actual and expected action's outcomes. Reward prediction error (RPE) signals have been observed in the midbrain, ventral striatum and ventromedial prefrontal cortex (vmPFC) (Schultz et al., 1997; O'Doherty et al., 2001, 2004; Pessiglione et al., 2006; D'Ardenne et al., 2008; Steinberg et al., 2013; Palminteri et al., 2015; Gueguen et al., 2021). Punishment prediction error (PPE) signals have been found in the anterior insula (aINS), dorsolateral prefrontal cortex (dlPFC), lateral orbitofrontal cortex (lOFC) and amygdala (O'Doherty et al., 2001; Seymour et al., 2005; Pessiglione et al., 2006; Yacubian et al., 2006; Gueguen et al., 2021). Finally, evidence from pharmacological manipulations and lesion studies indicate that reward and punishment learning can be selectively affected (Frank et al., 2004; Bódi et al., 2009; Palminteri et al., 2009, 2012). Complementary evidence, however, suggests that reward and punishment learning may instead share common neural substrates. Indeed, hubs of the reward circuit, such as the midbrain dopamine systems and vmPFC, contain neural populations encoding also punishments (Tom et al., 2007; Matsumoto and Hikosaka, 2009; Plassmann et al., 2010; Monosov and Hikosaka, 2012).

Yet, learning is known to reflect a network phenomenon emerging from neural interactions distributed over cortical-subcortical circuits (Bassett and Mattar, 2017; Hunt and Hayden, 2017; Averbeck and Murray, 2020; Averbeck and O'Doherty, 2022). More generally, cognitive functions emerge from the dynamic coordination over large-scale and hierarchically organized networks (Varela et al., 2001; Bressler and Menon, 2010; Reid et al., 2019; Panzeri et al., 2022; Thiebaut de Schotten and Forkel, 2022) and accumulating evidence supports that information about task variables is widely distributed across brain circuits, rather than anatomically localized (Parras et al., 2017; Saleem et al., 2018; Steinmetz et al., 2019; Urai et al., 2022; Voitov and Mrsic-Flogel, 2022).

Accordingly, we hypothesized that reward and punishment learning arise from complementary neural interactions between frontal cortex regions. In particular, we tested whether redundancy- or synergy-dominated interactions allow the emergence of collective brain networks differentially supporting reward and punishment learning. Indeed, redundancy-dominated brain networks can be associated with neural interactions subserving similar functions (i.e., functional segregation). Redundant interactions may appear in collective states dominated by oscillatory synchronization (Engel et al., 2001; Varela et al., 2001; Buzsáki and Draguhn, 2004; Fries, 2015) or resonance phenomena (Vinck et al., 2023). Such collective states may give rise to selective patterns of information flow (Buehlmann and Deco, 2010; Kirst et al., 2016; Palmigiano et al., 2017; Battaglia and Brovelli, 2020). Synergy-dominated brain networks can be associated with functionally-complementary interactions (i.e., functional

integration). Indeed, synergistic interactions have reported between distant transmodal regions during high-level cognition (Luppi et al., 2022) and, at the microscale, in populations of neurons within a cortical column of the visual cortex and across areas of the visuomotor network (Nigam et al., 2019; Varley et al., 2023).

We investigated neural interactions within and between four cortical regions, namely the aINS, dlPFC, lOFC and vmPFC, by means of intracerebral EEG (iEEG) data collected from epileptic patients while performing a reinforcement learning task (Pessiglione et al., 2006; Palminteri et al., 2012; Gueguen et al., 2021). We found various proportions of intracranial recordings encoding uniquely RPE or PPE signals or both, suggesting a local probabilistic representation of PEs. We then identified two distinct learning-related subsystems dominated by redundant interactions. A first subsystem with RPE-only interactions between the vmPFC and lOFC, and a second subsystem with PPE-only interactions between the aINS and dlPFC. Within each redundant-dominated subsystem, we demonstrated differential patterns of directional interactions, with the vmPFC and aINS playing a driving role in the reward and punishment learning circuits, respectively. Finally, these two subsystems interacted during the encoding of PE signals irrespectively of the context (reward or punishment), through synergistic collaboration between the dlPFC and vmPFC. We concluded that the functional specificity toward reward or punishment learning is better disentangled by interactions compared to local representations. Overall, our results provide a unifying explanation of distributed cortical representations and interactions supporting reward and punishment learning.

Results

iEEG data, behavioral task and computational modeling

We analyzed iEEG data from seventeen pharmacoresistant epileptic patients implanted with intracranial electrodes (Gueguen et al., 2021). A total of 248 iEEG bipolar derivations located in the aINS, dlPFC, vmPFC and lOFC regions (Fig. 1A) and 1788 pairs of iEEG signals, both within and across brain regions (Fig. 1B) were selected for further analysis. Participants performed a probabilistic instrumental learning task and had to choose between two cues to either maximize monetary gains (for reward cues) or minimize monetary losses (for punishment cues) (Fig. 1C). Overall, they selected more monetary gains and avoided monetary losses but the task structure was designed so that the number of trials was balanced between reward and punishment conditions (Fig. 1D).

We estimated trial-wise prediction errors by fitting a Q-learning model to behavioral data. Fitting the model consisted in adjusting the constant parameters to maximize the likelihood of observed choices. We used three constant parameters: i) the learning rate α accounting for how fast participants learned new pairs of cues; ii) the choice temperature β to model different levels of exploration and exploitation; iii) Θ parameter to account for the tendency to repeat the choice made in the previous trial. The RPE and PPE were obtained by taking the PE for rewarding and punishing pairs of cues, respectively. RPE and PPE showed high absolute values early during learning and tended toward zero as participants learned to predict the outcome (Fig. 1E).

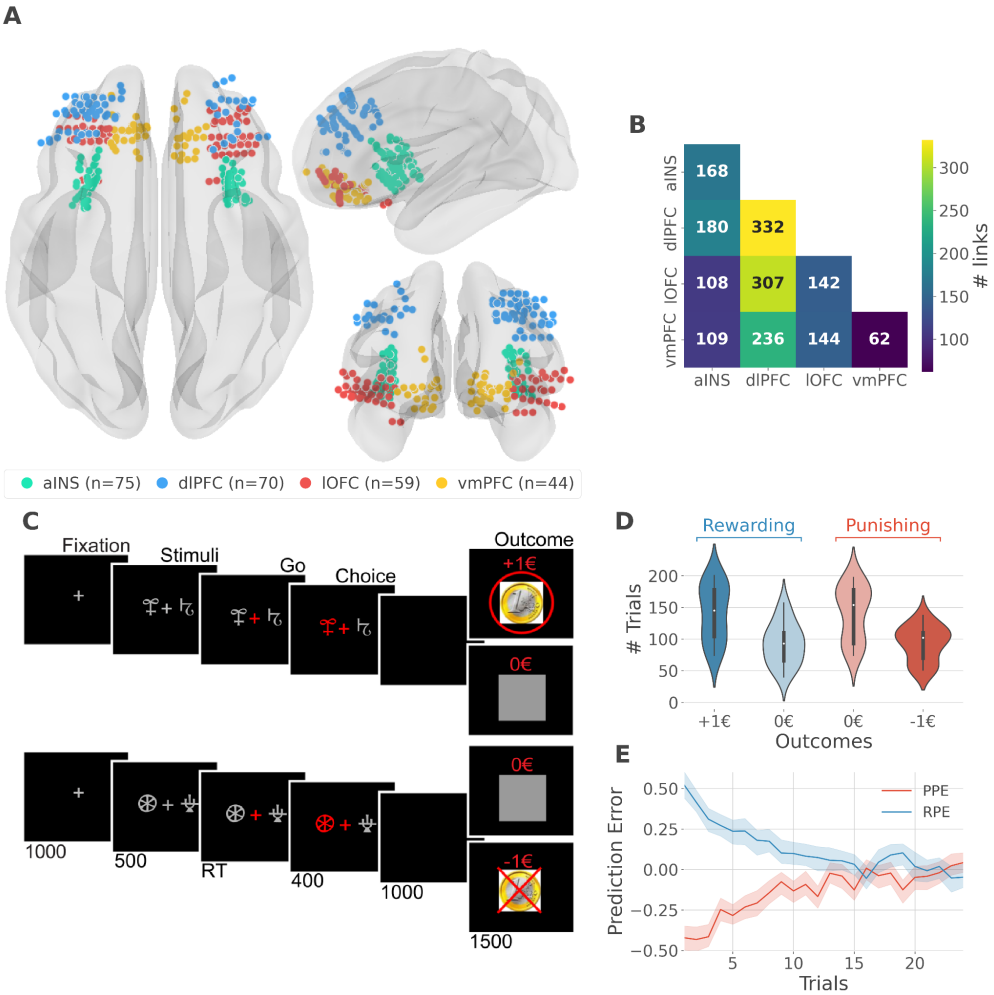

Fig. 1 iEEG implantation, behavioral task and computational modeling. *(A)* Anatomical location of intracerebral electrodes across the 17 epileptic patients. Anterior insula (aINS, n=75), dorsolateral prefrontal cortex (dlPFC, n=70), lateral orbitofrontal cortex (lOFC, n=59), ventromedial prefrontal cortex (vmPFC, n=44), *(B)* Number of pairwise connectivity links (i.e. within patients) within and across regions, *(C)* Example of a typical trial in the reward (top) and punishment (bottom) conditions. Participants had to select one abstract visual cue among the two presented on each side of a central visual fixation cross and subsequently observed the outcome. Duration is given in milliseconds, *(D)* Number of trials where participants received outcomes +1€ (142±44, mean±std) vs. 0€ (93±33) in the rewarding condition (blue) and outcomes 0€ (141±42) vs. -1€ (93±27) in the punishment condition (red), *(E)* Across participants trial-wise reward PE (RPE - blue) and punishment PE (PPE - red), ± 95% confidence interval.

Local probabilistic encoding of PE signals

At the neural level, we first investigated local correlates of prediction error signals by studying whether RPEs and PPEs are differentially encoded in prefrontal and insular regions. To this end, we performed model-based information theoretical analyses of iEEG gamma activities by computing the mutual information (MI) between the across-trials modulations in RPE or PPE signals and the gamma band power in the aINS, dlPFC, lOFC and vmPFC. Preliminary spectrally-resolved analyses showed that the frequency range significantly encoding prediction errors was between 50 and 100 Hz (Fig. S1). We thus extracted for each trial time-resolved gamma power within the 50 to 100 Hz range using a multi-taper approach for further analyses. MI analysis between gamma power and prediction error signals displayed significant group-level effects in all four cortical regions (Fig. 2A) and globally reproduced previous findings based on general linear model analyses (Gueguen et al., 2021). Interestingly, we observed a clear spatial dissociation between reward and punishment PE signaling. Whereas the vmPFC and dlPFC displayed complementary functional preferences for RPE and PPE, respectively, the aINS and the lOFC carried similar amount of information about both R/PPE (Figure 2A), although the strongest effect was observed in the aINS.

To better characterize the spatial granularity of PE encoding, we further studied the specificity of individual brain regions by categorizing contacts as either: i) RPE-specific; ii) PPE-specific; iii) PE-unspecific responding to both R/PPE; iv) PE-irrelevant (i.e. non-significant ones) (Fig. 2B). All regions displayed a local "probabilistic" encoding of prediction errors with temporal dynamics peaking around 500 ms after outcome presentation. The vmPFC and dlPFC differentially responded to reward and punishment PEs, and contained approximately 30% of RPE- and PPE-specific contacts, respectively. In both regions, the proportion of RPE- and PPE-specific contacts was elevated for approximately 1 s after outcome presentation. The lOFC also contained a large proportion of PPE-specific contacts, but displayed more transient dynamics lasting approximately 0.5 s. The aINS had similar proportions of iEEG sites specific for the RPE and PPE (20%), with temporal dynamics lasting approximately 0.75 s. Importantly, all regions contained approximately 10% of PE-unspecific iEEG sites that responded to both RPE and PPE, especially in the aINS and dlPFC. The remaining iEEG contacts were categorized as PE-irrelevant.

Taken together, our results demonstrate that reward and avoidance learning are not supported by highly selective brain activations, but rather from a mixed or probabilistic encoding of RPE or PPE signals distributed over the prefrontal and insular cortices. Nevertheless, such distributed encoding seems to involve two complementary systems primarily centered over the vmPC and dlPFC, respectively.

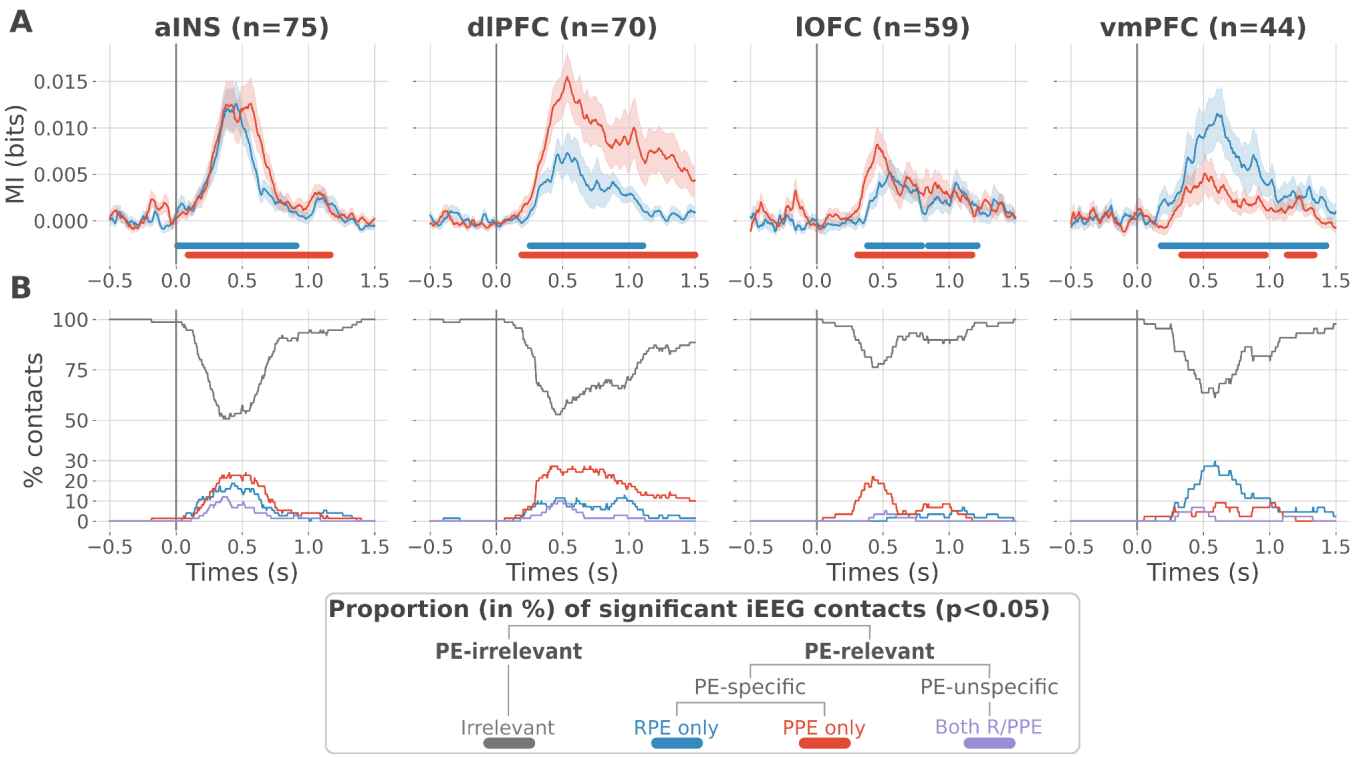

Fig. 2 Local probabilistic encoding of reward and punishment prediction error signals. *(A)* Time-courses of mutual information (MI in bits) estimated between the gamma power and the reward (blue) and punishment (red) PE signals. The solid line and the shaded area represent the mean and SEM of the across-contacts MI. Significant clusters of MI at the group level are plotted with horizontal bold lines ($p<0.05$, cluster-based correction, non-parametric randomization across epochs), *(B)* Instantaneous proportions of task-irrelevant (gray) and task-relevant iEEG recording contacts presenting a significant relation with either the RPE (blue), the PPE (red) or with both RPE and PPE (purple). Data are aligned to the outcome presentation (vertical line at 0 seconds).

Encoding of PE signals emerge from redundancy-dominated subsystems

To better understand the observed complex encoding of reward and punishment PEs, we tested the hypothesis that functional dissociations emerge from differential and distributed interactions between

prefrontal and insular cortices. To address this question, we performed model-based network-level analyses based on recent advances from information theory, the so-called Partial Information Decomposition (PID) framework (Williams and Beer, 2010; Wibral et al., 2017; Lizier et al., 2018). Within this framework, the interaction information (McGill, 1954; Ince et al., 2017), a multivariate generalization of mutual information, has been shown to quantify whether a three-variable interaction (i.e., pairs of brain regions and the PE variable) is either synergy- and redundancy-dominated (Williams and Beer, 2010). Interaction information (II), unlike mutual information, can be either positive or negative. A negative value indicates a net redundancy (i.e., pair of recordings are carrying similar information about the PE) interaction between variables, whereas a positive value indicates a net synergistic effect (i.e., pair of recordings are carrying complementary information about the PE).,

We computed the time-resolved II across trials between the gamma activity of pairs of iEEG signals and PEs. To differentiate cortico-cortical interactions for reward and punishment learning, we first calculated the II separately for RPEs and PPEs. RPE- and PPE-specific analyses exclusively showed negative modulations of II, therefore indicating the presence of redundancy-dominated interactions (Fig. 3). To better characterize the local interactions encoding reward and punishment PEs, we computed the II between pairs of gamma band signals recorded within the aINS, dlPFC, lOFC and vmPFC. Within-region II analyses showed that significant RPE-specific interactions were exclusively observed in the vmPFC and lOFC, whereas PPE-specific interactions were present only in the dlPFC. In addition, the aINS was found to display both RPE- and PPE-specific interactions (Fig. 3A). A relevant sign of high specificity for either reward or punishment PE signals was the presence of a significant cluster dissociating RPE and PPE in the vmPFC and dlPFC only (green clusters in Figure 3A).

To investigate the nature of long-range interactions, we computed the II for RPE and PPE between signals from different brain regions (Fig. 3B). Similarly, results exclusively showed redundancy-dominated interactions (i.e., negative modulations). RPE-specific interactions were observed between the lOFC and vmPFC, whereas PPE-specific interactions were observed between the aINS and dlPFC and to a smaller extent between the dlPFC and lOFC, peaking 500 ms after outcome presentation. A significant difference between RPE and PPE was exclusively observed in the lOFC-vmPFC and aINS-dlPFC interactions, but not between dlPFC and lOFC (green clusters in Fig. 3B). We conclude that the encoding of RPE and PPE signals emerge from redundancy-dominated subsystems that differentially engage prefronto-insular regions.

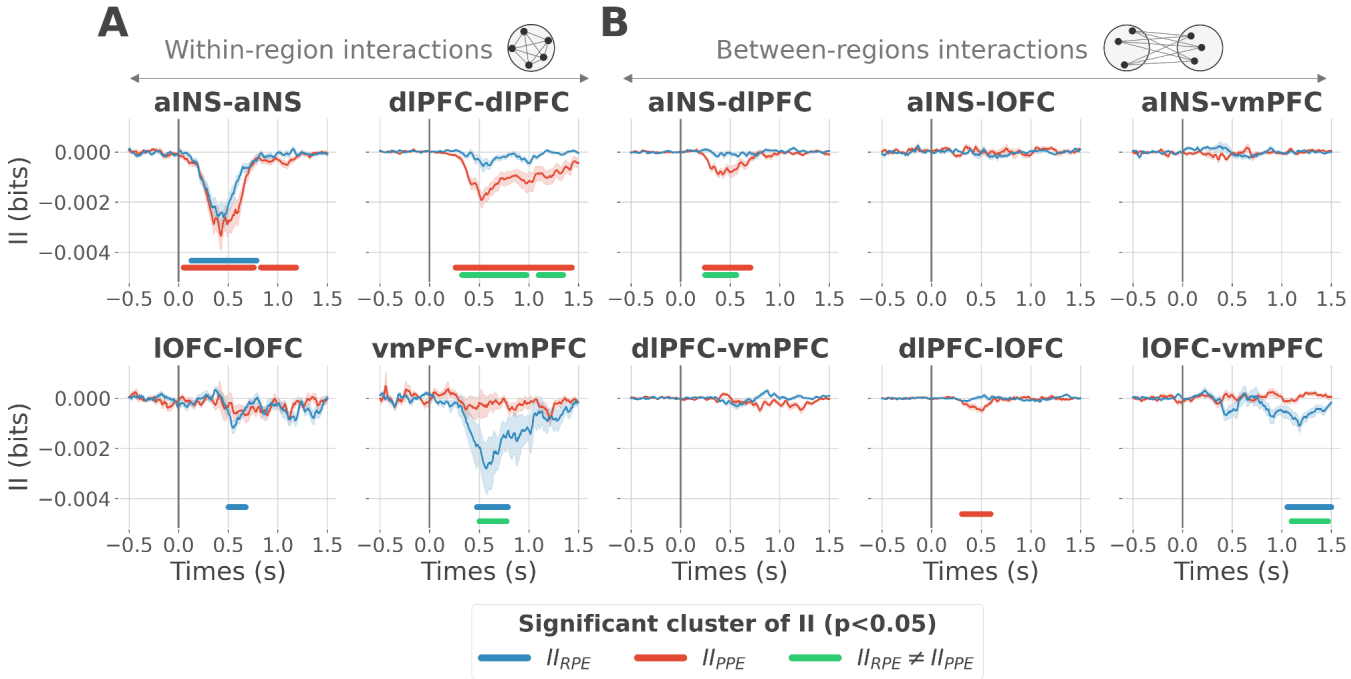

Fig. 3 Encoding of PE signals emerges from redundancy-dominated subsystems. Dynamic interaction information (II in bits) within- (*A*) and between-regions (*B*) about the RPE (II_{RPE}) and PPE (II_{PPE}) are plotted in blue and red. Significant clusters of II_{RPE} and II_{PPE} are displayed with horizontal bold blue and red lines ($p<0.05$, cluster-based correction, non-parametric randomization across epochs). Significant differences between II_{RPE} and II_{PPE} are displayed in green. Shaded areas represent the SEM. The vertical gray line at 0 seconds represents the outcome presentation.

Contextual directional interactions within redundant subsystems

Previous analyses of II are blind to the direction of information flows. To address this issue, we estimated the transfer entropy (TE) (Schreiber, 2000) on the gamma power during the rewarding (TE_{Rew}) and punishment conditions (TE_{Pun}), between all possible pairs of contacts. As a reminder, the TE is an information-theoretic measure that quantifies the degree of directed statistical dependence or "information flow" between time series, as defined by the Wiener-Granger principle (Wiener, 1956; Granger, 1969). Delay-specific analyses of TE showed that a maximum delay of information transfer between pairs of signals comprised an interval between 116 and 236 msec (Fig. S2). We thus computed the TE for all pairs of brain regions within this range of delays and detected temporal clusters where the TE significantly differed between conditions ($TE_{Rew} > TE_{Pun}$ or $TE_{Pun} > TE_{Rew}$). Only two pairs of brain regions displayed statistically-significant modulations in TE (Fig. 4). We observed that the TE from the aINS to the dlPFC ($TE_{aINS \rightarrow dlPFC}$) peaked at approximately 400 msec after outcome onset and was significantly stronger during the punishment condition compared to the rewarding condition. By contrast, the information flow around ~800ms from the vmPFC to the lOFC ($TE_{vmPFC \rightarrow lOFC}$) was significantly stronger during the rewarding condition. No other brain interactions were found significant (Fig. S3). Overall, these results demonstrate that the two redundancy-dominated RPE- and PPE-specific networks (Fig. 3B) are characterized by differential directional interactions. The vmPFC and aINS act as drivers in the two systems, whereas the dlPFC and lOFC play the role of receivers, thus suggesting a flow of PE-specific information within the network.

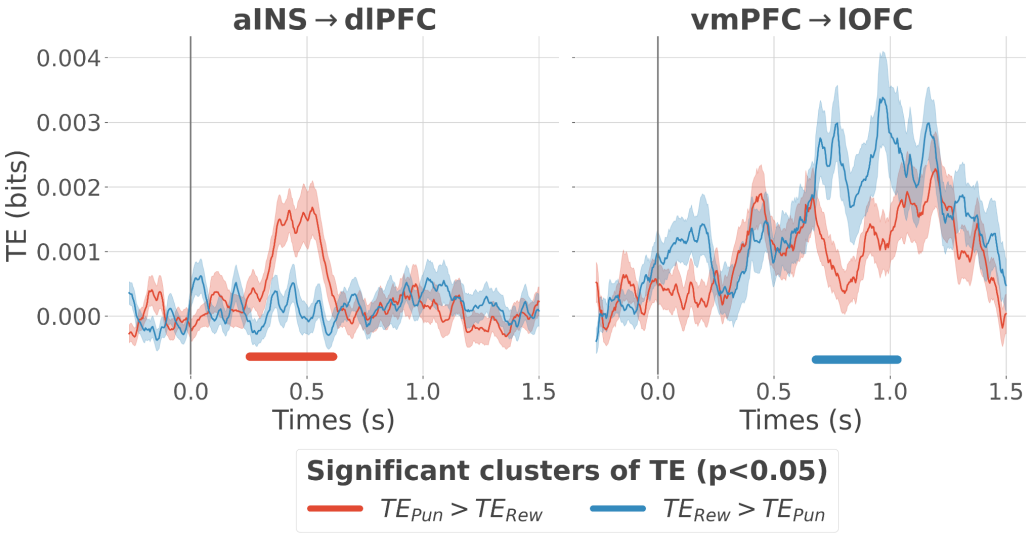

Fig. 4 Contextual modulation of information transfer. Time courses of transfer entropy (TE, in bits) from the aINS to the dlPFC (aINS→dlPFC) and from the vmPFC to the lOFC (vmPFC→lOFC), estimated during the rewarding condition (TE_{Rew} in blue) and punishing condition (TE_{Pun} in red). Significant differences ($p<0.05$, cluster-based correction, non-parametric randomization across epochs) of TE between conditions are displayed with horizontal bold lines (blue for $TE_{Rew} > TE_{Pun}$ and red for $TE_{Pun} > TE_{Rew}$). Shaded areas represent the SEM. The vertical gray line at 0 seconds represents the outcome presentation.

Integration of PE signals emerges from synergy-dominated interactions between segregated sub-systems

Since learning required participants to concurrently explore rewarding and punishment outcomes, we finally investigated the nature of cortico-cortical interactions encoding both RPE and PPE signals. We estimated the II about the full PEs, i.e. the information carried by co-modulation of gamma power between all pairs of contacts about PE signals (Fig S4). Encoding of PEs was specifically associated with significantly positive II between the dlPFC and vmPFC (Fig. 5A). Such between-regions synergy-dominated interaction occurred approximately between 250 and 600 ms after outcome onset.

We then investigated if the synergy between the dlPFC and vmPFC encoding global PEs could be explained by their respective local specificity. Indeed, we previously reported larger proportions of recordings encoding the PPE in the dlPFC and the RPE in the vmPFC (Fig. 2B). Therefore, it is possible that the $II_{dlPFC\text{-}vmPFC}$ is mainly due to complementary roles where the dlPFC brings information about the PPE only and the vmPFC brings information to the RPE only. To answer this question, we computed the $II_{dlPFC\text{-}vmPFC}$ for groups of bipolar derivations with different local specificity. As a reminder, bipolar derivations were previously categorized as RPE or PPE specific if their gamma activity were modulated according to the RPE only, to the PPE only or to both (Fig. 2B). We obtained four categories of II: the $II_{RPE\text{-}RPE}$ and $II_{PPE\text{-}PPE}$ reflecting the II estimated between RPE- and PPE-specific recordings, respectively, the $II_{PPE\text{-}RPE}$ between PPE- and RPE-specific contacts and the II_{Mixed} for the remaining possibilities (i.e., RPE-Both, PPE-Both and Both-Both) (Fig. S5A). Interestingly, we found significant synergistic interactions between recordings with mixed specificity i.e. $II_{PPE\text{-}RPE}$ and II_{Mixed} between 250 and 600 ms after outcome onset (Fig. 5B and 5C). Consequently, the $II_{dlPFC\text{-}vmPFC}$ is partly explained by the dlPFC and vmPFC carrying PPE- and RPE-specific information ($II_{PPE\text{-}RPE}$) together with interactions between non-specific recordings (II_{Mixed}). Taken together, the integration of the global PE signals emerged from the synergistic interaction between recordings with mixed specificity from the dlPFC and vmPFC.

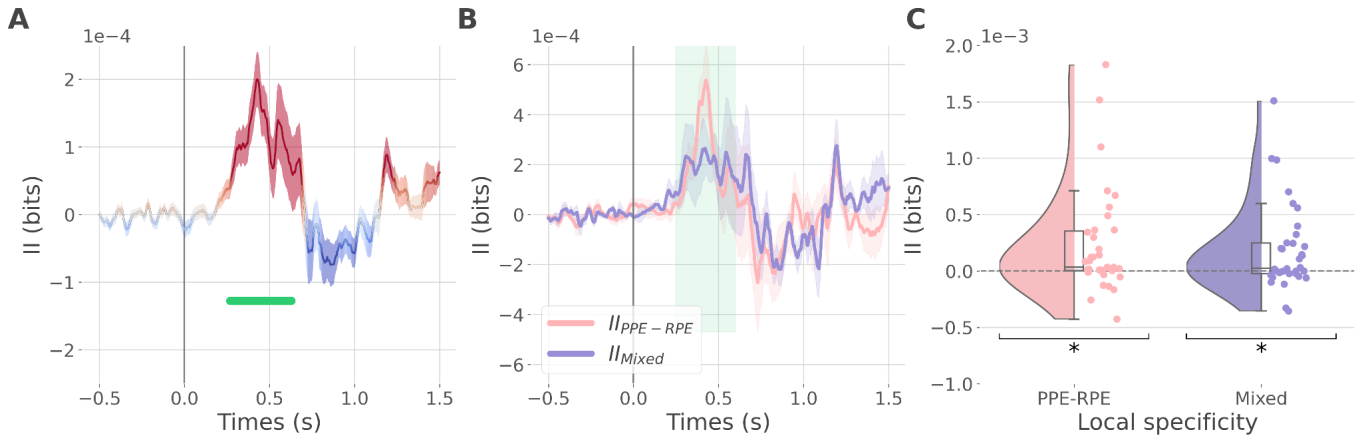

Fig 5. Synergistic interactions about the full PE signals between recordings of the dlPFC and vmPFC. *(A)* Dynamic interaction information (II in bits) between the dlPFC and vmPFC about the full prediction error ($II_{dlPFC\text{-}vmPFC}$). Hot and cold colors indicate synergy- and redundancy-dominated II about the full PE. Significant clusters of II are displayed with a horizontal bold green line (p<0.05, cluster-based correction, non-parametric randomization across epochs). Shaded areas represent the SEM. The vertical gray line at 0 seconds represents the outcome presentation. *(B)* Dynamic $II_{dlPFC\text{-}vmPFC}$ binned according to the local specificity PPE-RPE ($II_{PPE\text{-}RPE}$ in pink) or mixed (II_{Mixed} in purple) *(C)* Distributions of the mean of the $II_{PPE\text{-}RPE}$ and II_{Mixed} for each pair of recordings ($II_{PPE\text{-}RPE}$: one sample t-test against 0; dof=34; P fdr-corrected=0.015*; T=2.86; CI(95%)=[6.5e-5, 3.9e-4]; II_{Mixed}: dof=33; P fdr-corrected=0.015*; T=2.84; CI(95%)=[5.4e-5, 3.3e-4]).

Discussion

Our study revealed the presence of specific functional interactions between prefrontal and insular cortices about reward and punishment prediction error signals. We first provided evidence for a probabilistic encoding of reward and punishment prediction error signals in each cortical region. We then identified a first subsystem specifically encoding RPEs with emerging redundancy-dominated interactions within and between the vmPFC and lOFC, with a driving role of the vmPFC. A second subsystem specifically encoding PPEs emerged from redundancy-dominated interactions within and between the aINS and dlPFC, with a driving role of the aINS. Switching between the encoding of reward and punishment PEs involved a synergy-dominated interaction between these two systems mediated by interactions between the dlPFC and vmPFC (Fig. 6).

Local mixed representations of prediction errors. Amongst the four investigated core-learning regions, the vmPFC was the only region to show a higher group-level preference for RPEs. This supports the notion that the vmPFC is functionally more specialized for the processing outcomes in reward learning, as previously put forward by human fMRI meta-analyses (Yacubian et al., 2006; Diekhof et al., 2012; Bartra et al., 2013; Garrison et al., 2013; Fouragnan et al., 2018). The dlPFC, instead, showed a stronger selectivity for punishment PE, thus supporting results from fMRI studies showing selective activations for aversive outcomes (Liu et al., 2011; Garrison et al., 2013; Fouragnan et al., 2018). On the contrary, the aINS and lOFC did not show clear selectivity for either reward or punishment PEs. The aINS carried a comparable amount of information about the RPE and PPE, thus suggesting that the insula is part of the surprise-encoding network (Fouragnan et al., 2018; Loued-Khenissi et al., 2020). Previous study reported a stronger link between the gamma activity of the aINS and the PPE compared to the RPE (Gueguen et al., 2021). This discrepancy in the results could be explained by the removal of iEEG sites contaminated with pathological activity and by the measures of information we are using here that are able to detect both linear and non-linear relationships between gamma activity and PE signals (Ince et al., 2017). The lOFC showed an initial temporal selectivity for PPE followed by a delayed one about the RPE. This is in accordance with fMRI and human intracranial studies which revealed that the lOFC was activated when receiving punishing outcomes, but also contains reward-related information (O'Doherty et al., 2001; Saez et al., 2018; Gueguen et al., 2021).

By taking advantage of the multi-site sampling of iEEG recordings, we quantified the heterogeneity in functional selectivity within each area and showed that the region-specific tendency toward either RPE or PPEs (Fig. 2A) could be explained by the largest domain-specific proportion of contacts (Fig. 2B). In other words, if a region showed a larger proportion of contacts being RPE-specific, the amount of information about the RPE at the group-level was also larger. Interestingly, we observed that 5 to 20% of contacts within a given region encoded both the RPE and PPE, thus revealing local mixed representations. Consequently, a strict dichotomous classification of learning-related areas as either reward and punishment may fail to capture important properties of the individual nodes of the learning circuit, such as the functional heterogeneity in the encoding of PEs. These results suggest that the human prefrontal cortex exhibits a mixed local selectivity for prediction error signals at the mesoscopic scale. This view is in line with recent literature showing that prefrontal cortex contains single neurons exhibiting mixed selectivity for multiple task variables (Meyers et al., 2008; Rigotti et al., 2013; Stokes et al., 2013; Panzeri et al., 2015; Parthasarathy et al., 2017; Bernardi et al., 2020). In the learning domain, single-unit studies have reported neurons encoding both rewarding and aversive outcomes in the OFC of the primate (Morrison and Salzman, 2009; Monosov and Hikosaka, 2012; Hirokawa et al., 2019). Mixed selectivity provides computational benefits, such as increasing the number of binary classifications, improving cognitive flexibility, and simplifying readout by downstream neurons (Fusi et al., 2016; Helfrich and Knight, 2019; Ohnuki et al., 2021; Panzeri et al., 2022). We suggest that the encoding of cognitive variables such as prediction error signals is supported by similar principles based on mixed selectivity at the meso- and macroscopic level, and may provide a natural substrate for cognitive flexibility and goal-directed learning (Rigotti et al., 2013).

Redundancy-dominated interactions segregate reward and punishment learning subsystems. We then tested whether the encoding of RPE and PPE signals could be supported by differential cortico-cortical interactions within and between frontal brain regions. To do so, we exploited the interaction information (II) (McGill, 1954; Ince et al., 2017) to quantify whether the amount of information bound up in a pair of gamma responses and PE signals is dominated by redundant or synergistic interactions (Williams and Beer, 2010). The II revealed redundancy-dominated interactions specific for RPE and PPE in the vmPFC and the dlPFC, respectively (Fig. 3A). The aINS was the only region for which the between-contacts II did not increase the functional selectivity, with large redundant interactions for both RPE and PPE signals. This suggests that within-area redundant interactions can potentially amplify the functional specificity, despite the presence of local mixed selectivity (Fig. 2A). Such “winner-take-all” competition could be implemented by mutual inhibition mechanisms, which have been suggested to be essential in reward-guided choice (Hunt et al., 2012; Jocham et al., 2012; Strait et al., 2014; Hunt and Hayden, 2017).

Across-areas interaction information revealed two subsystems with intrinsic redundancy-dominated interactions. A reward subsystem with RPE-specific interactions between the lOFC and vmPFC, and a punishment subsystem with PPE-specific interactions between the aINS and dlPFC (Fig. 3B). Although a significant modulation selective for RPE was also present in the interaction between dlPFC and lOFC peaking around 500 ms after outcome presentation, a significant difference between the encoding of RPE and PPE was exclusively observed in the lOFC-vmPFC and aINS-dlPFC interactions (green clusters in Fig. 3B). This result suggests that the observed functionally-distinct learning circuits for RPE and PPEs emerge from differential cortico-cortical interactions, rather than distinct local properties. More generally, our results suggest that redundancy-based network-level interactions are related with the functional

specificity observed in neuroimaging and lesion studies (Pessiglione and Delgado, 2015; Palminteri and Pessiglione, 2017).

We then investigated differential communication patterns and directional relations within the two redundancy-dominated circuits (Kirst et al., 2016; Palmigiano et al., 2017). We identified significant information routing patterns, dissociating reward and punishment learning (Fig. 4). Within the reward subsystem, the vmPFC played a driving role toward the lOFC only during the rewarding condition. Conversely, within the punishment subsystem, the aINS played a driving role toward the dlPFC only during the punishment condition. These results support the notion that redundancy-dominated cognitive networks are associated with the emergence of information-routing capabilities, where signals are communicated on top of collective reference states (Battaglia and Brovelli, 2020).

Encoding the full PE is supported by synergistic interactions between subsystems. Humans can flexibly switch between learning strategies that allow the acquisition of stimulus-action-outcomes associations in changing contexts. We investigated how RPE and PPE subsystems coordinated to allow such behavioral flexibility. To do so, we searched for neural correlates of PEs irrespectively of the context (reward or punishment learning) in between-regions interactions. We found that the encoding of global PE signals was associated with synergy-dominated interactions between the two subsystems, mediated by the interactions between the dlPFC and the vmPFC (Fig. 5). Importantly, such synergy-dominated interaction reveals that the joint representation of the dlPFC and vmPFC is greater than the sum of their individual contributions for the encoding of global PE signals. Thus, it suggests that successful adaptation in varying contexts requires both the vmPFC and dlPFC for the encoding of global PE signals.

Role of redundant and synergistic interactions in brain network coordination. At the macroscopic level, few studies investigated the potential role of redundant and synergistic interactions. By combining functional and diffusion MRI, recent work suggested that redundant interactions are predominantly associated with structurally coupled and functionally segregated processing. In contrast, synergistic interactions preferentially support functional integrative processes and complex cognition across higher-order brain networks (Luppi et al., 2022). Triadic synergistic interactions between the continuous spike counts recorded within and across areas of the visuomotor network have been shown to carry behaviorally-relevant information and to display the strongest modulations during the processing of visual information and movement execution (Varley et al., 2023). Finally, cortical representations of prediction error signals in the acoustic domain observed tone-related and instantaneous redundant interactions, such as time-lagged synergistic interactions within and across temporal and frontal regions of the auditory system (Gelens et al., 2023).

At the microscopic level, the amount of information encoded by a population of neurons can be modulated by pairwise and higher-order interactions, producing varying fractions of redundancy and synergy (Averbeck et al., 2006; Panzeri et al., 2015, 2022). Synergistic and redundant pairs of neurons can be identified by estimating the amount of information contained in the joint representation minus the sum of the information carried by individual neurons (Schneidman et al., 2003). Redundant coding is intricately linked to correlated activity (Gutnisky and Dragoi, 2008) and can spontaneously emerge due to the spatial correlations present in natural scenes by triggering neurons with overlapping receptive fields. Correlations between the trial-by-trial variations of neuronal responses could limit the amount of information encoded by a population (Ramon Bartolo et al., 2020; Kafashan et al., 2021) and facilitate readout by downstream neurons (Salinas and Sejnowski, 2001). While redundancy has been at the heart of heated debates and influential theories, such as efficient coding and redundancy compression in

sensory areas (Barlow, 2001), synergy phenomena have been described to a lesser extent. Recently, a study reported synergistic coding in a V1 cortical column together with structured correlations between synergistic and redundant hubs (Nigam et al., 2019). Taken together, the literature associated with population codes proposed that balancing proportions of redundancy and synergy offers a good compromise between system robustness and resilience to cell loss and the creation of new information (Panzeri et al., 2022).

Taken together, we suggest that redundancy-dominated interactions confer robustness and network-level selectivity for complementary learning processes, which may lead to functional integration processes. On the other hand, synergy-dominated interactions seem to support neural interactions between redundancy-dominated networks, thus supporting functional integrative processes in the brain. In addition, our study suggests that redundant and synergistic interactions occur across multiple spatial scales from local to large-scale.

Conclusion. Our report of mixed representation of reward and punishment prediction error signals explains the discrepancy in the attribution of a functional specificity to the core learning cortical regions. Instead, we propose that functional specialization for reward and punishment PE signals emerges from redundancy-dominated interactions within the two subsystems formed by the vmPFC-lOFC and aINS-dlPFC, respectively. Within each subsystem, we observed asymmetric and directional interactions with the vmPFC and aINS playing a driving role in the reward and punishment learning circuits. Finally, switching between reward and punishment learning was supported by synergistic collaboration between subsystems. This supports the idea that higher-order integration between functionally-distinct subsystems are mediated by synergistic interactions. Taken together, our results provide a unifying view reconciling distributed cortical representations with interactions supporting reward and punishment learning. They highlight the relevance of considering learning as a network-level phenomenon by linking distributed and functionally redundant subnetworks through synergistic interactions hence supporting flexible cognition (Fedorenko and Thompson-Schill, 2014; Petersen and Sporns, 2015; Bassett and Mattar, 2017; Hunt and Hayden, 2017; Averbeck and Murray, 2020; Averbeck and O'Doherty, 2022).

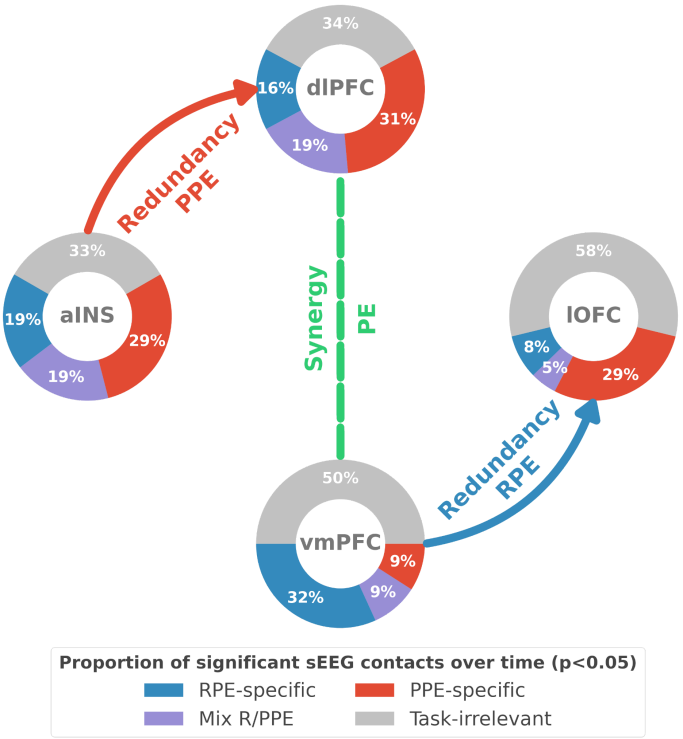

Fig 6. Summary of findings. The four nodes represent the investigated regions, namely the anterior insula (aINS), the dorsolateral and ventromedial parts of the prefrontal cortex (dlPFC and vmPFC) and the lateral orbitofrontal cortex (lOFC). The outer disc represents the local probabilistic encoding i.e. the different proportions of contacts over time having a significant relationship between the gamma power and PE signals. In blue, is the proportion of contacts with a significant relation with the PE across rewarding trials (RPE-specific). Conversely, in red for punishment trials (PPE-specific). In purple, the proportion of contacts with a significant relationship with both the RPE and PPE. In gray, is the remaining proportion of non-significant contacts. Regarding interactions, we found that information transfer between aINS and dlPFC carried redundant information about PPE only and information transfer between vmPFC and lOFC about RPE only. This information transfer occurred with a leading role of the aINS in the punishment context and the vmPFC in the rewarding context. Finally, we found synergistic interactions between the dlPFC and the vmPFC about the full PE, without splitting into rewarding and punishing conditions.

Methods

Data acquisition and experimental procedure

Intracranial EEG recordings

Intracranial electroencephalography (iEEG) recordings were collected from seventeen patients presenting pharmaco-resistant focal epilepsy and undergoing presurgical evaluation (33.5 ± 12.4 years old, 10 females). As the location of the epileptic foci could not be identified through noninvasive methods, neural activity was monitored using intracranial stereotactic electroencephalography. Multi-lead and semi-rigid depth electrodes were stereotactically implanted according to the suspected origin of seizures. The selection of implantation sites was based solely on clinical aspects. iEEG recordings were performed at the clinical neurophysiology epilepsy departments of Grenoble and Lyon Hospitals (France). iEEG electrodes had a diameter of 0.8mm, 2mm wide, 1.5mm apart and contained 8–18 contact leads (Dixi, Besançon, France). For each patient, 5 to 17 electrodes were implanted. Recordings were conducted using an audio–video-EEG monitoring system (Micromed, Treviso, Italy), which allowed simultaneous recording of depth iEEG channels sampled at 512 Hz (6 patients), or 1024 Hz (12 patients) [0.1–200 Hz bandwidth]. One of the contacts located in the white matter was used as a reference. Anatomical localizations of iEEG contacts were determined based on post-implant computed tomography scans or post-implant MRI scans coregistered with pre-implantation scans (Lachaux et al., 2003; Mercier et al.,

2022). All patients gave written informed consent and the study received approval from the ethics committee (CPP 09-CHUG-12, study 0907) and from a competent authority (ANSM no: 2009-A00239-48).

Limitations

iEEG have been collected from pharmacoresistant epileptic patients who underwent deep electrode probing for preoperative evaluation. However, we interpreted these data as if collected from healthy subjects and assumed that epileptic activity does not affect the neural realization of prediction error. To best address this question, we excluded electrodes contaminated with pathological activity and focused on task-related changes and multi-trial analysis to reduce the impact of incorrect or task-independent neural activations. Therefore, our results may benefit from future replication in healthy controls using non-invasive recordings. Despite the aforementioned limitations, we believe that access to deep intracerebral EEG recordings of human subjects can provide privileged insight into the neural dynamics that regulate human cognition, with outstanding spatial, temporal, and spectral precision. In the long run, this type of data could help bridge the gap between neuroimaging studies and electrophysiological recordings in nonhuman primates.

Preprocessing of iEEG data

Bipolar derivations were computed between adjacent electrode contacts to diminish contributions of distant electric sources through volume conduction, reduce artifacts and increase the spatial specificity of the neural data. Bipolar iEEG signals can approximately be considered as originating from a cortical volume centered within two contacts (Brovelli et al., 2005; Bastin et al., 2016; Combrisson et al., 2017), thus providing a spatial resolution of approximately 1.5-3mm (Lachaux et al., 2003; Jerbi et al., 2009; Mercier et al., 2022). Recording sites with artifacts and pathological activity (e.g. epileptic spikes) were removed using visual inspection of all of the traces of each site and each participant.

Definition of anatomical regions of interest

Anatomical labeling of bipolar derivations was performed using the IntrAnat software (Deman et al., 2018). The 3D T1 pre-implantation MRI gray/white matter was segmented and spatially normalized to obtain a series of cortical parcels using MarsAtlas (Auzias et al., 2016) and the Destrieux atlas (Destrieux et al., 2010). 3D coordinates of electrodes contacts were then coregistered on post-implantation images (MRI or CT). Each recording site (i.e., bipolar derivation) was labeled according to its position in a parcellation scheme in the participant's native space. Thus, the analyzed dataset only included electrodes identified to be in the gray matter. Four regions of interest (ROIs) were defined for further analysis: 1) the ventromedial prefrontal cortex (vmPFC) ROI was created by merging six (three per hemisphere) parcels in MarsAlas (labeled PFCvm, OFCv and OFCvm in MarsAtlas) corresponding to the ventromedial prefrontal cortex and fronto-medial part of the orbitofrontal cortex, respectively; 2) the lateral orbitofrontal cortex (lOFC) ROI included four (two per hemisphere) MarsAtlas parcels (MarsAtlas labels: OFCvl and the OFCv); 3) the dorsolateral prefrontal cortex (dlPFC) ROI was defined as the inferior and superior bilateral dorsal prefrontal cortex (MarsAtlas labels: PFrdli and PFrdls); 4) the anterior insula (aINS) ROI was defined as the bilateral anterior part of the insula (Destrieux atlas labels: Short insular gyri, anterior circular insular sulcus and anterior portion of the superior circular insular sulcus). The total number of bipolar iEEG derivations for the four ROIS was 44, 59, 70 and 75 for the vmPFC, lOFC, dlPFC and aINS, respectively (Fig. 1A).

Behavioral task and set-up

Participants were asked to participate in a probabilistic instrumental learning task adapted from previous studies (Pessiglione et al., 2006; Palminteri et al., 2012). Participants received written instructions that the goal of the task was to maximize their financial payoff by considering reward-seeking and punishment avoidance as equally important. Instructions were reformulated orally if necessary. Participants started with a short session, with only two pairs of cues presented on 16 trials, followed by 2–3 short sessions of 5 min. At the end of this short training, all participants were familiar with the timing of events, with the response buttons and all reached a threshold of at least 70% of correct choices during both reward and punishment conditions. Participants then performed three to six sessions on a single testing occurrence, with short breaks between sessions. Each session was an independent task, with four new pairs of cues to be learned. Cues were abstract visual stimuli taken from the Agathodaimon alphabet. The two cues of a pair were always presented together on the left and right of a central fixation cross and their relative position was counterbalanced across trials. On each trial, one pair was randomly presented. Each pair of cues was presented 24 times for a total of 96 trials per session. The four pairs of cues were divided into two conditions. A rewarding condition where the two pairs could either lead the participants to win one euro or nothing (+1€ vs. 0€) and a symmetric punishment condition where the participants could either lose one euro or nothing (-1€ vs. 0€). Rewarding and punishing pairs of cues were presented in an intermingled random manner and participants had to learn the four pairs at once. Within each pair, the two cues were associated with the two possible outcomes with reciprocal probabilities (0.75/0.25 and 0.25/0.75). To choose between the left or right cues, participants used their left or right index to press the corresponding button on a joystick (Logitech Dual Action). Since the position on the screen was counterbalanced, response (left versus right) and value (good vs. bad cue) were orthogonal. The chosen cue was colored in red for 250 ms and then the outcome was displayed on the screen after 1000 ms. To win money, participants had to learn by trial and error which cue-outcome association was the most rewarding in the rewarding condition and the least penalizing in the punishment condition. Visual stimuli were delivered on a 19-inch TFT monitor with a refresh rate of 60 Hz, controlled by a PC with Presentation 16.5 (Neurobehavioral Systems, Albany, CA).

Computational model of learning

To model choice behavior and estimate prediction error signals, we used a standard Q-learning model (Watkins and Dayan, 1992) from reinforcement learning theory (Sutton and Barto, 2018). For a pair of cues *A* and *B*, the model estimates the expected value of choosing *A* (*Qa*) or *B* (*Qb*), given previous choices and received outcomes. Q-values were initiated to 0, corresponding to the average of all possible outcome values. After each trial *t*, the expected value of choosing a stimulus (e.g. *A*) was updated according to the following update rule:

$$Qa_{t+1} = Qa_t + \alpha\delta_t \quad \textbf{(1)}$$

with α the learning rate weighting the importance given to new experiences and δ, the outcome prediction error (PE) signals at a trial *t* defined as the difference between the obtained and expected outcomes:

$$\delta_t = R_t - Qa_t \quad \textbf{(2)}$$

with R_t the reinforcement value among -1€, 0€ and 1€. The probability of choosing a cue was then estimated by transforming the expected values associated with each cue using a softmax rule with a Gibbs distribution. An additional Θ parameter was added in the softmax function to the expected value of the

chosen option on the previous trial of the same cue to account for the tendency to repeat the choice made on the previous trial. For example, if a participant chose option *A* on trial *t*, the probability of choosing *A* at trial *t+1* was obtained using :

$$Pa_{t+1} = \frac{e^{(Qa_t+\theta)/\beta}}{e^{(Qa_t+\theta)/\beta} + e^{Qb_t/\beta}} \quad \textbf{(3)}$$

with β the choice temperature for controlling the ratio between exploration and exploitation. The three free parameters α, β and Θ were fitted per participant and optimized by minimizing the negative log-likelihood of choice using the MATLAB *fmincon* function, initialized at multiple starting points of the parameter space (Palminteri et al., 2015). Estimates of the free parameters, the goodness of fit and the comparison between modeled and observed data can be seen in Table 1 and Figure 1 in (Gueguen et al., 2021).

iEEG data analysis

Estimate of single-trial gamma-band activity

A key neural signal for studying functional cortical networks is the gamma-band (40–150 Hz) neural activity. In humans, power modulations in the gamma range are commonly recorded using MEG and intracranial EEG to map task-related brain regions (Brovelli et al., 2005; Crone et al., 2006; Vidal et al., 2006; Ball et al., 2008; Jerbi et al., 2009; Lachaux et al., 2012; Cheyne and Ferrari, 2013). In addition, gamma-band modulations can be used to characterize cortico-cortical interactions and functional connectivity among brain regions supporting executive functions (Brovelli et al., 2015, 2017; Combrisson et al., 2022a). In the current study, we estimated the power in the gamma band using a multitaper time-frequency transform based on Slepian tapers (Percival et al., 1993; Mitra and Pesaran, 1999). To extract gamma-band activity from 50 to 100Hz, the iEEG time series were multiplied by 9 orthogonal tapers (15 cycles for a duration of 200ms and with a time-bandwidth for frequency smoothing of 10Hz), centered at 75 Hz and Fourier-transformed. To limit false negatives proportions due to multiple testings, we down-sampled the gamma power to 256 Hz. Finally, we smoothed the gamma power using a 10 points Savitzky-Golay filter. We used MNE-Python (Gramfort et al., 2013) to inspect the time series, reject contacts contaminated with pathological activity, and estimate the power spectrum density (mne.time_frequency.psd_multitaper) and the gamma power (mne.time_frequency.tfr_multitaper).

Local correlates of PE signals

To quantify the local encoding of prediction error (PE) signals in the four ROIs, we used information-theoretic metrics. To this end, we computed the time-resolved mutual information (MI) between the single-trial gamma-band responses and the outcome-related PE signals. As a reminder, mutual information is defined as:

$$I(X;Y) = H(X) - H(X|Y) \quad \textbf{(4)}$$

In this equation, the variables X and Y represent the across-trials gamma-band power and the PE variables, respectively. H(X) is the entropy of X, and H(X|Y) is the conditional entropy of X given Y. In the current study, we used a semi-parametric binning-free technique to calculate MI, called Gaussian-Copula Mutual Information (GCMI) (Ince et al., 2017). The GCMI is a robust rank-based

approach that allows the detection of any type of monotonic relation between variables and it has been successfully applied to brain signals analysis (Colenbier et al., 2020; Michelmann et al., 2021; Ten Oever et al., 2021). Mathematically, the GCMI is a lower-bound estimation of the true MI and it does not depend on the marginal distributions of the variables, but only on the copula function that encapsulates their dependence. The rank-based copula-normalization preserves the relationship between variables as long as this relation is strictly increasing or decreasing. As a consequence, the GCMI can only detect monotonic relationships. Nevertheless, the GCMI is of practical importance for brain signal analysis for several reasons. It allows to estimate the MI on a limited number of samples and it contains a parametric bias correction to compensate for the bias due to the estimation on smaller datasets. It allows to compute the MI on uni- and multivariate variables that can either be continuous or discrete (see Table I in (Ince et al., 2017)). Finally, it is computationally efficient, which is a desired property when dealing with a large number of iEEG contacts recording at a high sampling rate. Here, the GCMI was computed across trials and it was used to estimate the instantaneous amount of information shared between the gamma power of iEEG contacts and RPE ($MI_{RPE} = I(\gamma; RPE)$) and PPE signals ($M_{PPE} = I(\gamma; PPE)$).

Network-level interactions and PE signals

The goal of network-level analyses was to characterize the nature of cortico-cortical interactions encoding reward and punishment PE signals. In particular, we aimed to quantify: 1) the nature of the interdependence between pairs of brain ROIs in the encoding of PE signals; 2) the information flow between ROIs encoding PE signals. These two questions were addressed using Interaction Information and Transfer Entropy analyses, respectively.

Interaction Information analysis: In classical information theory, interaction information (II) provides a generalization of mutual information for more than two variables (McGill, 1954; Ince et al., 2017). For the three-variables case, the II can be defined as the difference between the total, or joint, mutual information between ROIs (R_1 and R_2) and the third behavioral variable (S), minus the two individual mutual information between each ROI and the behavioral variable. For a three variables multivariate system composed of two sources R_1, R_2 and a target S, the II is defined as :

$$\begin{aligned} II(R_1; R_2; S) &= I(S; R_1|R_2) - I(R_1; S) \\ &= I(R_1, R_2; S) - I(R_1; S) - I(R_2; S) \end{aligned} \quad \textbf{(5)}$$

Unlike mutual information, the interaction information can be either positive or negative. A negative value of interaction information indicates a net redundant effect between variables, whereas positive values indicate a net synergistic effect (Williams and Beer, 2010). Here, we used the II to investigate the amount of information and the nature of the interactions between the gamma power of pairs of contacts (γ_1, γ_2) about the RPE ($II_{RPE} = II(\gamma_1, \gamma_2; RPE)$) and PPE signals ($II_{PPE} = II(\gamma_1, \gamma_2; PPE)$). The II was computed by estimating the MI quantities of equation *(5)* using the GCMI between contacts within the same brain region or across different regions.

Transfer Entropy analysis: To quantify the degree of communication between neural signals, the most successful model-free methods rely on the Wiener-Granger principle (Wiener, 1956; Granger, 1969). This principle identifies information flow between time series when future values of a given signal can be predicted from the past values of another, above and beyond what can be achieved from its autocorrelation. One of the most general information theoretic measures based on the Wiener-Granger

principle is Transfer Entropy (TE) (Schreiber, 2000). The TE can be formulated in terms of conditional mutual information (Schreiber, 2000; Kaiser and Schreiber, 2002):

$$TE(X \rightarrow Y) = I(X_{past}; Y_t | Y_{past}) \quad \textbf{(6)}$$

Here, we computed the TE on the gamma activity time courses of pairs of iEEG contacts. We used the GCMI to estimate conditional mutual information. For an interval [d_1, d_2] of n_{delays}, the final TE estimation was defined as the mean over the TE estimated at each delay :

$$TE(X \rightarrow Y)_{[d_1,d_2]} = \frac{1}{n_{delays}} . \sum_{d=d_1}^{d_2} I(X_d; Y_t | Y_d) \quad \textbf{(7)}$$

Statistical analysis

We used a group-level approach based on non-parametric permutations, encompassing non-negative measures of information (Combrisson et al., 2022a). The same framework was used at the local level (i.e. the information carried by single contact) or at the network level (i.e. the information carried by pairs of contacts for the II and TE). To take into account the inherent variability existing at the local and network levels, we used a random-effect model. To generate the distribution of permutations at the local level, we shuffled the PE variable across trials 1000 times and computed the MI between the gamma power and the shuffled version of the PE. The shuffling led to a distribution of MI reachable by chance, for each contact and at each time point (Combrisson and Jerbi, 2015). To form the group-level effect, we computed a one-sample t-test against the permutation mean across the MI computed on individual contacts taken from the same brain region, at each time point. The same procedure was used on the permutation distribution to form the group-level effect reachable by chance. We used cluster-based statistics for correcting for multiple comparisons (Maris and Oostenveld, 2007). The cluster-forming threshold was defined as the 95th percentile of the distribution of t-values obtained from the permutations. We used this threshold to form the temporal clusters within each brain region. We obtained cluster masses on both the true t-values and the t-values computed on the permutations. To correct for multiple comparisons, we built a distribution made of the largest 1000 cluster masses estimated on the permuted data. The final corrected p-values were inferred as the proportion of permutations exceeding the t-values. To generate the distributions of II and TE reachable by chance, we respectively shuffled the PE variable across trials for the II and the gamma power across trials of the source for the TE (Vicente et al., 2011). The rest of the significance testing procedure at the network level is similar to the local level, except that it is not applied within brain regions but within pairs of brain regions. Finally, for information-theoretic metrics and group-level statistics, we used a homemade Python software called *Frites* (Combrisson et al., 2022b).

Acknowledgments

EC, RB, JB and AB were supported by the PRC project “CausaL” (ANR-18-CE28-0016). This project/research has received funding from the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). AB were supported by FLAG ERA II “Joint Transnational Call 2017" - HBP - Basic and Applied Research 2, Brainsynch-Hit (ANR-17-HBPR-0001). RB acknowledges support through a PhD Scholarship awarded by the Neuroschool. This work has received support from the French government under the Programme Investissements d’Avenir, Initiative d’Excellence d’Aix-Marseille Université via

A*Midex (AMX-19-IET-004) and ANR (ANR-17-EURE-0029) funding. JB was supported by ANR-17-CE37-0018 and ANR-18-CE28-0016. Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high-performance computing resources.

References

Auzias G, Coulon O, Brovelli A (2016) MarsAtlas : A cortical parcellation atlas for functional mapping: MarsAtlas. Hum Brain Mapp 37:1573–1592.

Averbeck B, O'Doherty JP (2022) Reinforcement-learning in fronto-striatal circuits. Neuropsychopharmacology 47:147–162.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366.

Averbeck BB, Murray EA (2020) Hypothalamic Interactions with Large-Scale Neural Circuits Underlying Reinforcement Learning and Motivated Behavior. Trends Neurosci 43:681–694.

Ball T, Demandt E, Mutschler I, Neitzel E, Mehring C, Vogt K, Aertsen A, Schulze-Bonhage A (2008) Movement related activity in the high gamma range of the human EEG. Neuroimage 41:302–310.

Balleine BW (2019) The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. Neuron 104:47–62.

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419.

Balleine BW, O'doherty JP (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol 35:48–69.

Barlow H (2001) Redundancy reduction revisited. Netw Comput Neural Syst 12:241.

Bartra O, McGuire JT, Kable JW (2013) The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage 76:412–427.

Bassett DS, Mattar MG (2017) A Network Neuroscience of Human Learning: Potential to Inform Quantitative Theories of Brain and Behavior. Trends Cogn Sci 21:250–264.

Bastin J, Deman P, David O, Gueguen M, Benis D, Minotti L, Hoffman D, Combrisson E, Kujala J, Perrone-Bertolotti M, Kahane P, Lachaux J-P, Jerbi K (2016) Direct Recordings from Human Anterior Insula Reveal its Leading Role within the Error-Monitoring Network. Cereb Cortex:bhv352.

Battaglia D, Brovelli A (2020) Functional connectivity and neuronal dynamics: insights from computational methods. Cogn Neurosci Sixth Ed.

Bernardi S, Benna MK, Rigotti M, Munuera J, Fusi S, Salzman CD (2020) The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. Cell 183:954-967.e21.

Bódi N, Kéri S, Nagy H, Moustafa A, Myers CE, Daw N, Dibó G, Takats A, Bereczki D, Gluck MA (2009) Reward-learning and the novelty-seeking personality: a between-and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain J Neurol 132:2385–2395.

Bouton ME (2007) Learning and behavior: A contemporary synthesis.

Bressler SL, Menon V (2010) Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn Sci 14:277–290.

Brovelli A, Badier J-M, Bonini F, Bartolomei F, Coulon O, Auzias G (2017) Dynamic Reconfiguration of Visuomotor-Related Functional Connectivity Networks. J Neurosci 37:839–853.

Brovelli A, Lachaux J-P, Kahane P, Boussaoud D (2005) High gamma frequency oscillatory activity dissociates attention from intention in the human premotor cortex. NeuroImage 28:154–164.

Brovelli XA, Chicharro D, Badier J-M, Wang H, Jirsa V (2015) Characterization of Cortical Networks and Corticocortical Functional Connectivity Mediating Arbitrary Visuomotor Mapping. J Neurosci 35:12643–12658.

Buehlmann A, Deco G (2010) Optimal Information Transfer in the Cortex through Synchronization Friston KJ, ed. PLoS Comput Biol 6:e1000934.

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. science 304:1926–1929.

Cheyne D, Ferrari P (2013) MEG studies of motor cortex gamma oscillations: evidence for a gamma "fingerprint" in the brain? Front Hum Neurosci 7:575.

Colenbier N, Van de Steen F, Uddin LQ, Poldrack RA, Calhoun VD, Marinazzo D (2020) Disambiguating the role of blood flow and global signal with partial information decomposition. Neuroimage 213:116699.

Combrisson E, Allegra M, Basanisi R, Ince RAA, Giordano BL, Bastin J, Brovelli A (2022a) Group-level inference of information-based measures for the analyses of cognitive brain networks from neurophysiological data. NeuroImage 258:119347.

Combrisson E, Basanisi R, Cordeiro VL, Ince RAA, Brovelli A (2022b) Frites: A Python package for functional connectivity analysis and group-level statistics of neurophysiological data. J Open Source Softw 7:3842.

Combrisson E, Jerbi K (2015) Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods 250:126–136.

Combrisson E, Perrone-Bertolotti M, Soto JL, Alamian G, Kahane P, Lachaux J-P, Guillot A, Jerbi K (2017) From intentions to actions: Neural oscillations encode motor processes through phase, amplitude and phase-amplitude coupling. NeuroImage 147:473–487.

Crone NE, Sinai A, Korzeniewska A (2006) High-frequency gamma oscillations and human brain mapping with electrocorticography. In: Progress in Brain Research, pp 275–295. Elsevier. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0079612306590193 [Accessed June 15, 2016].

D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319:1264–1267.

Deman P, Bhattacharjee M, Tadel F, Job A-S, Rivière D, Cointepas Y, Kahane P, David O (2018) IntrAnat electrodes: a free database and visualization software for intracranial electroencephalographic data processed for case and group studies. Front Neuroinformatics 12:40.

Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53:1–15.

Dickinson A, Balleine B (1994) Motivational control of goal-directed action. Anim Learn Behav 22:1–18.

Diekhof EK, Kaps L, Falkai P, Gruber O (2012) The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude–An activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. Neuropsychologia 50:1252–1266.

Dolan RJ, Dayan P (2013) Goals and habits in the brain. Neuron 80:312–325.

Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top–down processing. Nat Rev Neurosci 2:704–716.

Fedorenko E, Thompson-Schill SL (2014) Reworking the language network. Trends Cogn Sci 18:120–126.

Fouragnan E, Retzler C, Philiastides MG (2018) Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. Hum Brain Mapp 39:2887–2906.

Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943.

Fries P (2015) Rhythms for Cognition: Communication through Coherence. Neuron 88:220–235.

Fusi S, Miller EK, Rigotti M (2016) Why neurons mix: high dimensionality for higher cognition. Curr Opin Neurobiol 37:66–74.

Garrison J, Erdeniz B, Done J (2013) Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310.

Gelens F, Komatsu M, Uran C, Jensen MA, Miller KJ, Ince RA, Vinck M, Canales-Johnson A (2023) Distributed representations of prediction error signals across the cortical hierarchy are synergistic. BioRxiv Prepr Serv Biol:2023–01.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, others (2013) MEG and EEG data analysis with MNE-Python. Front Neurosci 7:267.

Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. Econom J Econom Soc:424–438.

Gueguen MCM, Lopez-Persem A, Billeke P, Lachaux J-P, Rheims S, Kahane P, Minotti L, David O, Pessiglione M, Bastin J (2021) Anatomical dissociation of intracerebral signals for reward and punishment prediction errors in humans. Nat Commun 12:3344.

Gutnisky DA, Dragoi V (2008) Adaptive coding of visual information in neural populations. Nature 452:220–224.

Helfrich RF, Knight RT (2019) Cognitive neurophysiology of the prefrontal cortex. In: Handbook of Clinical Neurology, pp 35–59. Elsevier. Available at: https://linkinghub.elsevier.com/retrieve/pii/B9780128042816000033 [Accessed April 8, 2022].

Hirokawa J, Vaughan A, Masset P, Ott T, Kepecs A (2019) Frontal cortex neuron types categorically encode single decision variables. Nature 576:446–451.

Hunt LT, Hayden BY (2017) A distributed, hierarchical and recurrent

framework for reward-based choice. Nat Rev Neurosci 18:172–182.
Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MF, Behrens TE (2012) Mechanisms underlying cortical activity during value-guided choice. Nat Neurosci 15:470–476.
Ince RAA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG (2017) A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula: Gaussian Copula Mutual Information. Hum Brain Mapp 38:1541–1573.
Jerbi K, Ossandon T, Hamamé CM, Senova S, Dalal SS, Jung J, Minotti L, Bertrand O, Berthoz A, Kahane P, Lachaux J-P (2009) Task-related gamma-band dynamics from an intracerebral perspective: Review and implications for surface EEG and MEG. Hum Brain Mapp 30:1758–1771.
Jocham G, Hunt LT, Near J, Behrens TE (2012) A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. Nat Neurosci 15:960–961.
Kafashan M, Jaffe AW, Chettih SN, Nogueira R, Arandia-Romero I, Harvey CD, Moreno-Bote R, Drugowitsch J (2021) Scaling of sensory information in large neural populations shows signatures of information-limiting correlations. Nat Commun 12:473.
Kaiser A, Schreiber T (2002) Information transfer in continuous processes. Phys Nonlinear Phenom 166:43–62.
Kirst C, Timme M, Battaglia D (2016) Dynamic information routing in complex networks. Nat Commun 7:11061.
Lachaux J-P, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol 98:279–301.
Lachaux JP, Rudrauf D, Kahane P (2003) Intracranial EEG and human brain mapping. J Physiol-Paris 97:613–628.
Liu X, Hairston J, Schrier M, Fan J (2011) Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. Neurosci Biobehav Rev 35:1219–1236.
Lizier JT, Bertschinger N, Jost J, Wibral M (2018) Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work.
Loued-Khenissi L, Pfeuffer A, Einhäuser W, Preuschoff K (2020) Anterior insula reflects surprise in value-based decision-making and perception. NeuroImage 210:116549.
Luppi AI, Mediano PAM, Rosas FE, Holland N, Fryer TD, O'Brien JT, Rowe JB, Menon DK, Bor D, Stamatakis EA (2022) A synergistic core for human brain evolution and cognition. Nat Neurosci 25:771–782.
Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.
Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459:837–841.
McGill W (1954) Multivariate information transmission. Trans IRE Prof Group Inf Theory 4:93–111.
Mercier MR, Dubarry A-S, Tadel F, Avanzini P, Axmacher N, Cellier D, Del Vecchio M, Hamilton LS, Hermes D, Kahana MJ, others (2022) Advances in human intracranial electroencephalography research, guidelines and good practices. NeuroImage:119438.
Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T (2008) Dynamic population coding of category information in inferior temporal and prefrontal cortex. J Neurophysiol 100:1407–1419.
Michelmann S, Price AR, Aubrey B, Strauss CK, Doyle WK, Friedman D, Dugan PC, Devinsky O, Devore S, Flinker A (2021) Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. Nat Commun 12:1–15.
Mitra PP, Pesaran B (1999) Analysis of dynamic brain imaging data. Biophys J 76:691–708.
Monosov IE, Hikosaka O (2012) Regionally Distinct Processing of Rewards and Punishments by the Primate Ventromedial Prefrontal Cortex. J Neurosci 32:10318–10330.
Morrison SE, Salzman CD (2009) The Convergence of Information about Rewarding and Aversive Stimuli in Single Neurons. J Neurosci 29:11471–11483.
Nigam S, Pojoga S, Dragoi V (2019) Synergistic Coding of Visual Information in Columnar Networks. Neuron 104:402-411.e4.
O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. science 304:452–454.
O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 4:95–102.
Ohnuki T, Osako Y, Manabe H, Sakurai Y, Hirokawa J (2021) Over-representation of fundamental decision variables in the prefrontal cortex underlies decision bias. Neurosci Res 173:1–13.
Palmigiano A, Geisel T, Wolf F, Battaglia D (2017) Flexible information routing by transient synchrony. Nat Neurosci 20:1014–1022.
Palminteri S, Justo D, Jauffret C, Pavlicek B, Dauta A, Delmaire C, Czernecki V, Karachi C, Capelle L, Durr A, Pessiglione M (2012) Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. Neuron 76:998–1009.
Palminteri S, Khamassi M, Joffily M, Coricelli G (2015) Contextual modulation of value signals in reward and punishment learning. Nat Commun 6 Available at: http://www.nature.com/articles/ncomms9096 [Accessed January 2, 2019].
Palminteri S, Lebreton M, Worbe Y, Grabli D, Hartmann A, Pessiglione M (2009) Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. Proc Natl Acad Sci 106:19179–19184.
Palminteri S, Pessiglione M (2017) Opponent Brain Systems for Reward and Punishment Learning. In: Decision Neuroscience, pp 291–303. Elsevier. Available at: https://linkinghub.elsevier.com/retrieve/pii/B9780128053089000233 [Accessed February 7, 2019].
Panzeri S, Macke JH, Gross J, Kayser C (2015) Neural population coding: combining insights from microscopic and mass signals. Trends Cogn Sci 19:162–172.
Panzeri S, Moroni M, Safaai H, Harvey CD (2022) The structures and functions of correlations in neural population codes. Nat Rev Neurosci 23:551–567.
Parras GG, Nieto-Diego J, Carbajal GV, Valdés-Baizabal C, Escera C, Malmierca MS (2017) Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nat Commun 8:2148.
Parthasarathy A, Herikstad R, Bong JH, Medina FS, Libedinsky C, Yen S-C (2017) Mixed selectivity morphs population codes in prefrontal cortex. Nat Neurosci 20:1770–1779.
Percival DB, Walden AT, others (1993) Spectral analysis for physical applications. cambridge university press.
Pessiglione M, Delgado MR (2015) The good, the bad and the brain: neural correlates of appetitive and aversive values underlying decision making. Curr Opin Behav Sci 5:78–84.
Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442:1042–1045.
Petersen SE, Sporns O (2015) Brain networks and cognitive architectures. Neuron 88:207–219.
Plassmann H, O'Doherty JP, Rangel A (2010) Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. J Neurosci 30:10799–10808.
Ramon Bartolo, Bartolo R, Saunders RC, Mitz AR, Averbeck BB (2020) Information-Limiting Correlations in Large Neural Populations. J Neurosci 40:1668–1678.
Reid AT, Headley DB, Mill RD, Sanchez-Romero R, Uddin LQ, Marinazzo D, Lurie DJ, Valdés-Sosa PA, Hanson SJ, Biswal BB, Calhoun V, Poldrack RA, Cole MW (2019) Advancing functional connectivity research from association to causation. Nat Neurosci 22:1751–1760.
Rescorla RA, Wagner AR, Black AH, Prokasy WF (1972) Classical conditioning II: current research and theory.
Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. Nature 497:585–590.
Saez I, Lin J, Stolk A, Chang E, Parvizi J, Schalk G, Knight RT, Hsu M (2018) Encoding of Multiple Reward-Related Computations in Transient and Sustained High-Frequency Activity in Human OFC. Curr Biol 28:2889-2899.e3.
Saleem AB, Diamanti EM, Fournier J, Harris KD, Carandini M (2018) Coherent encoding of subjective spatial position in visual cortex and hippocampus. Nature 562:124–127.
Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. Nat Rev Neurosci 2:539–550.
Schneidman E, Bialek W, Berry MJ (2003) Synergy, Redundancy, and Independence in Population Codes. J Neurosci 23:11539–11553.
Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85:461.
Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction

and reward. Science 275:1593–1599.
Seymour B, O'doherty JP, Koltzenburg M, Wiech K, Frackowiak R, Friston K, Dolan R (2005) Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat Neurosci 8:1234–1240.
Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966–973.
Steinmetz NA, Zatka-Haas P, Carandini M, Harris KD (2019) Distributed coding of choice, action and engagement across the mouse brain. Nature 576:266–273.
Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. Neuron 78:364–375.
Strait CE, Blanchard TC, Hayden BY (2014) Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. Neuron 82:1357–1366.
Sutton RS, Barto AG (2018) Reinforcement learning: An introduction. MIT press.
Ten Oever S, Sack AT, Oehrn CR, Axmacher N (2021) An engram of intentionally forgotten information. Nat Commun 12:1–14.
Thiebaut de Schotten M, Forkel SJ (2022) The emergent properties of the connected brain. Science 378:505–510.
Thorndike EL (1898) Animal intelligence: An experimental study of the associative processes in animals. Psychol Rev Monogr Suppl 2:i.
Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The Neural Basis of Loss Aversion in Decision-Making Under Risk. Science 315:515–518.
Urai AE, Doiron B, Leifer AM, Churchland AK (2022) Large-scale neural recordings call for new insights to link brain and behavior. Nat Neurosci 25:11–19.
Varela F, Lachaux J-P, Rodriguez E, Martinerie J (2001) The brainweb: Phase synchronization and large-scale integration. Nat Rev Neurosci 2:229–239.
Varley TF, Sporns O, Schaffelhofer S, Scherberger H, Dann B (2023) Information-processing dynamics in neural networks of macaque cerebral cortex reflect cognitive state and behavior. Proc Natl Acad Sci 120:e2207677120.
Vicente R, Wibral M, Lindner M, Pipa G (2011) Transfer entropy—a model-free measure of effective connectivity for the neurosciences. J Comput Neurosci 30:45–67.
Vidal JR, Chaumon M, O'Regan JK, Tallon-Baudry C (2006) Visual grouping and the focusing of attention induce gamma-band oscillations at different frequencies in human magnetoencephalogram signals. J Cogn Neurosci 18:1850–1862.
Vinck M, Uran C, Spyropoulos G, Onorato I, Broggini AC, Schneider M, Canales-Johnson A (2023) Principles of large-scale neural interactions. Neuron 111:987–1002.
Voitov I, Mrsic-Flogel TD (2022) Cortical feedback loops bind distributed representations of working memory. Nature 608:381–389.
Watkins CJCH, Dayan P (1992) Q-learning. Mach Learn 8:279–292.
Wibral M, Priesemann V, Kay JW, Lizier JT, Phillips WA (2017) Partial information decomposition as a unified approach to the specification of neural goal functions. Brain Cogn 112:25–38.
Wiener N (1956) The theory of prediction. Mod Math Eng.
Williams PL, Beer RD (2010) Nonnegative Decomposition of Multivariate Information. ArXiv10042515 Math-Ph Physicsphysics Q-Bio Available at: http://arxiv.org/abs/1004.2515 [Accessed May 2, 2019].
Yacubian J, Gläscher J, Schroeder K, Sommer T, Braus DF, Büchel C (2006) Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. J Neurosci 26:9530–9537.

Supplementary

Local encoding of prediction error signals within the gamma band

We searched for the most informative sub-band about the R/PPE within the broad gamma range. To this end, we estimated the PSD between [50, 200]Hz during the first second following the outcome apparition. We used multitapers with a 20Hz bandwidth for the multi-taper windowing function. We then estimated the amount of information carried by individual gamma frequencies about the R/PPE in the four brain regions. We found significant clusters of MI over the four brain regions with both RPE and PPE signals mostly within the [50, 100]Hz range (Fig. S1A). We then estimated the density of information within the [50, 100]Hz, [100, 150]Hz and [150, 200]Hz bins (Fig. S1B). Approximately 60% of the total information was concentrated within the [50, 100]Hz range.

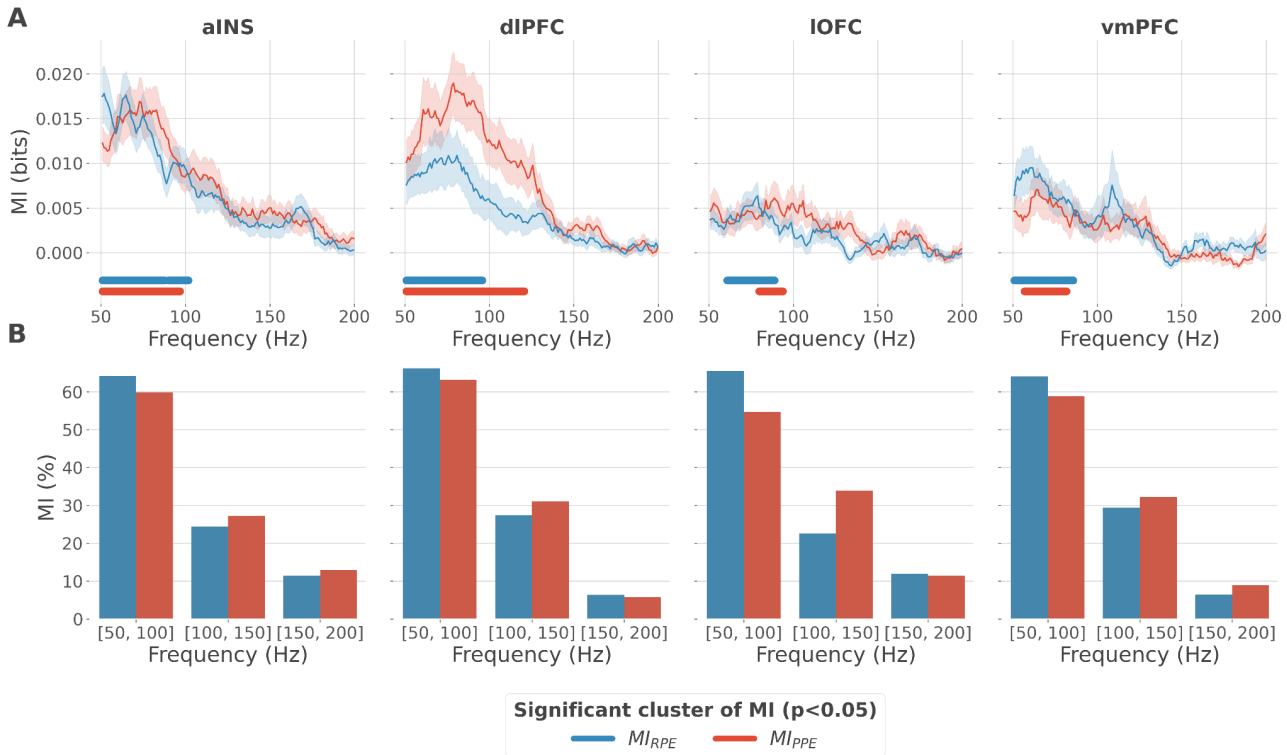

Fig S1. Local encoding of prediction error signals within the gamma band. *(A)* Distribution of information in the aINS, dlPFC, lOFC and vmPFC about the R/PPE in the frequency domain. The solid line and the shaded area respectively represent the mean and SEM of the across-contacts MI. Horizontal thick lines represent significant clusters of information ($p<0.05$, cluster-based correction, non-parametric randomization across epochs), *(B)* Density of information in the [50, 100]Hz, [100, 150]Hz and [150, 200]Hz bins.

Optimal delay interval for maximizing information transfer

We searched for the optimal delay interval to maximize the information transfer between regions. An important parameter to consider when estimating TE is the number of time points in the past of the target to use for conditioning. An information flow from source X to target Y exists because the inclusion of the past of X reduces the uncertainty about the future of Y, given its own past. We investigated the influence of the delay on the TE. To this end, we estimated the TE across all pairs of contacts per participant, for rewarding and punishing trials, at every possible delay up to 350ms (Fig. S2). We found a maximum information flow for delays up to 176 ms. Therefore, for the main text analyses, we used a range centered around 176±60ms ([116, 236]ms).

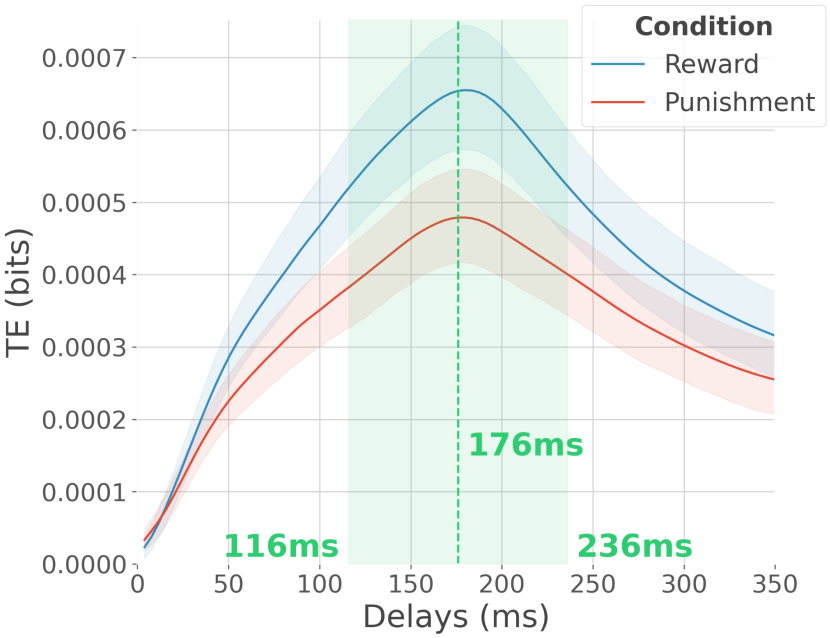

Fig S2. Optimal delay interval for maximizing information transfer. Modulation of transfer entropy (TE in bits) as a function of the delay between source and target areas. In blue, the TE computed across rewarding trials and in red, the TE computed across punishing trials. Shaded areas surrounding the time courses represent the 95% confidence interval estimated using a bootstrapping strategy.

Differential cortico-cortical directional interactions

We estimated the transfer entropy (TE) on the gamma power during the rewarding (TE_{Rew}) and punishment conditions (TE_{Pun}). As in the main text, we computed the TE for all pairs of brain regions within delays between 116 and 236 msec and detected temporal clusters where the TE significantly differed between conditions ($TE_{Rew} > TE_{Pun}$ or $TE_{Pun} > TE_{Rew}$). Only two pairs of brain regions displayed statistically-significant modulations in TE (Fig. S3). The TE from the aINS to the dlPFC ($TE_{aINS \rightarrow dlPFC}$) during the punishment condition and the TE from the vmPFC to the lOFC ($TE_{vmPFC \rightarrow lOFC}$) during the rewarding condition. No other brain interactions were found significant.

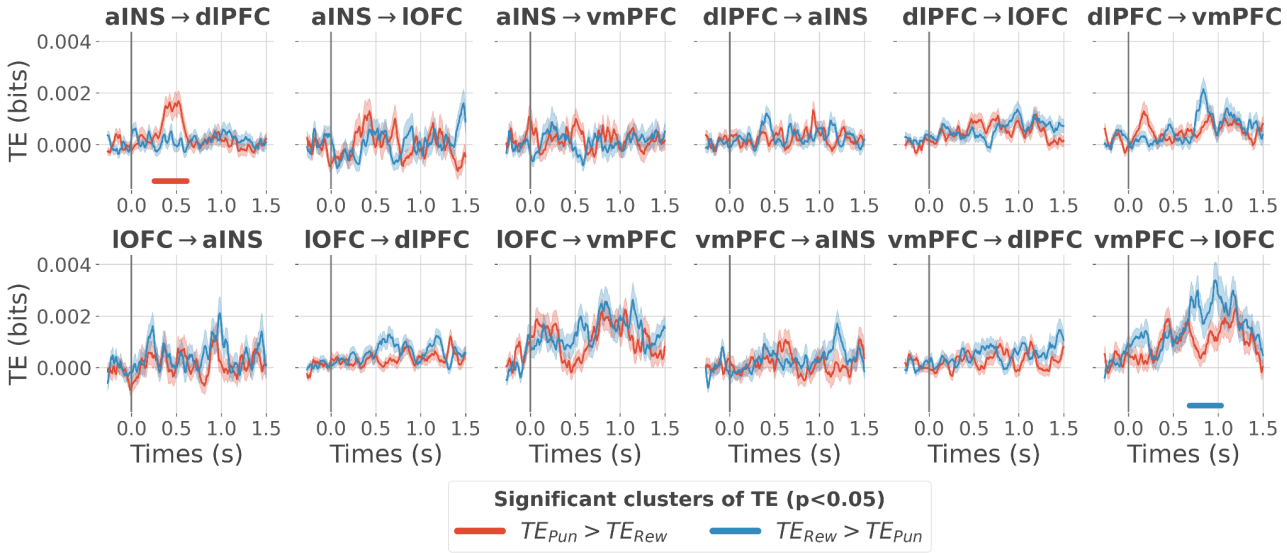

Fig S3. Contextual modulation of the information transfer. Time courses of transfer entropy (TE, in bits) estimated during the rewarding condition (TE_{Rew} in blue) and punishing condition (TE_{Pun} in red). Significant differences ($p<0.05$, cluster-based correction, non-parametric randomization across epochs) of TE between conditions are displayed with horizontal bold lines (blue for $TE_{Rew} > TE_{Pun}$ and red for $TE_{Pun} > TE_{Rew}$). Shaded areas represent the SEM. The vertical grey line at 0 seconds represents the outcome presentation.

Interactions about the full PE signals

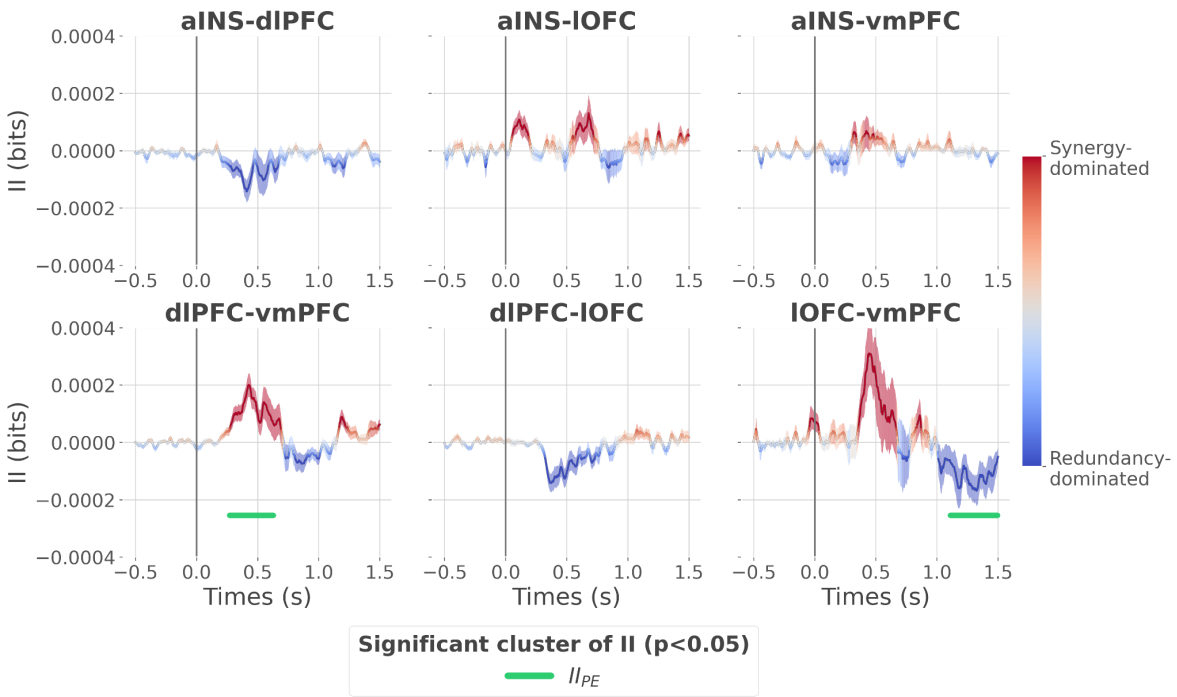

Fig S4. Cortico-cortical interactions about the full PE signals. Dynamic interaction information (II in bits) between-regions about the full prediction error (II_{PE}). Hot and cold colors indicate synergy- and redundancy-dominated interactions about the full PE. Significant clusters of II_{PE} are displayed with a horizontal bold green line ($p<0.05$, cluster-based correction, non-parametric randomization across epochs). Shaded areas represent the SEM. The vertical gray line at 0 seconds represents the outcome presentation.

Interaction information binned according to the local specificity

We binned the II about the full PE (i.e. by concatenating the RPE and PPE) according to the local specificity of the bipolar derivations in the dlPFC and vmPFC i.e. contacts with gamma activity modulated according to the RPE only, to the PPE only or to both RPE and PPE (Fig. 2B). As a result, we binned the II into four categories: the $II_{RPE\text{-}RPE}$ and $II_{PPE\text{-}PPE}$ respectively reflecting the II estimated between recordings specific to the RPE and PPE, the $II_{PPE\text{-}RPE}$ between recordings PPE and RPE specific and the II_{Mixed} for the remaining possibilities (i.e. RPE-Both, PPE-Both and Both-Both) (Fig. S5A). We reported a significant cluster II about the full PE approximately between [250; 600]ms (Fig. 5). Therefore, we estimated the mean II across time points between [250; 600]ms within each of the four categories, for each pair of recordings (Fig. S5B) and computed a one-sample t-test against 0 (Table 1). Only the $II_{PPE\text{-}RPE}$ (pink) and II_{Mixed} (purple) mean showed a significant difference from 0. However, the number of pairs of recordings for the $II_{RPE\text{-}RPE}$ (red, 7 pairs) and $II_{PPE\text{-}PPE}$ (blue, 6 pairs) were probably too small to find a significant difference.

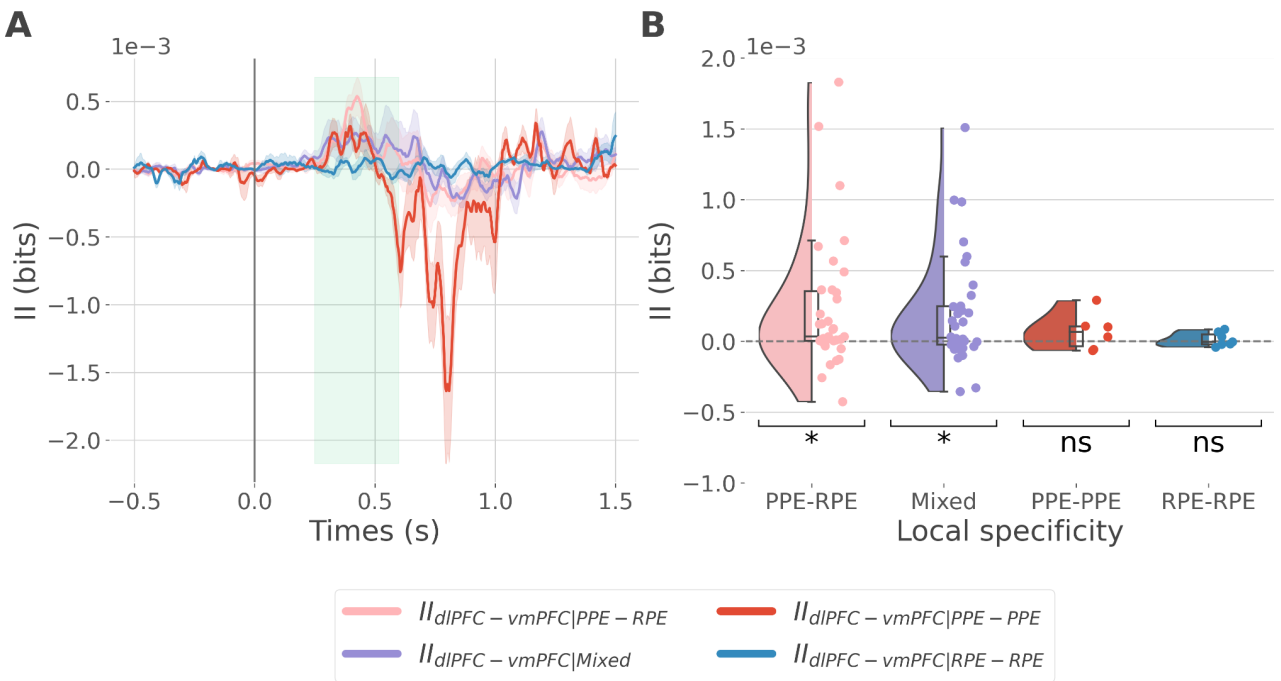

Fig S5. Interaction information binned according to the local specificity. *(A)* Dynamic interaction information (II in bits) between the dlPFC and vmPFC ($II_{dlPFC\text{-}vmPFC}$) binned according to the local specificity toward the RPE and PPE. Shaded areas represent the SEM. The vertical gray line at 0 seconds represents the outcome presentation. *(B)* Mean II between time points from 250 to 600ms after outcome presentation per category of local specificity. Each individual point represents one pairs of recordings from the dlPFC and vmPFC.

	T-value	P-value	P-value (FDR corrected)	dof	CI95%
$II_{PPE\text{-}RPE}$	2,859	0.007**	0.015*	34	[6.5e-05, 3.9e-04]
II_{Mixed}	2,841	0.008**	0.015*	33	[5.4e-05, 3.3e-04]
$II_{PPE\text{-}PPE}$	1,25	0.2667	0.3556	5	[-7.1e-05, 2.1e-04]
$II_{RPE\text{-}RPE}$	0,733	0.4912	0.4912	6	[-3.1e-05, 5.8e-05]

Table 1. Results of the one sample t-test performed against 0