Decoding Arbitrary and Informed Decisions from Intracranial Recordings in Humans

Laura Marras[1,9*], Maxime Verwoert[1], Maarten C. Ottenhoff[1], Sophocles Goulis[1], Johannes P. van Dijk[2,3,4], Simon Tousseyn[2], Louis Wagner[2], Albert J. Colon[2], Pieter L. Kubben[1,5], Marcus L.F. Janssen[6,7], Steffen A. Herff[8], and Christian Herff[1,*]

[1]Department of Neurosurgery, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands
[2]Academic Center for Epileptology, Kempenhaeghe/Maastricht University Medical Center, Heeze, The Netherlands
[3]Department of Orthodontics, Ulm University, Ulm, Germany
[4]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[5]Academic Center for Epileptology, Kempenhaeghe/Maastricht University Medical Center, Maastricht, The Netherlands
[6]Department of Clinical Neurophysiology, Maastricht University Medical Center, Maastricht, The Netherlands
[7]School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands
[8]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
[9]MoMiLab, IMT School for Advanced Studies Lucca, Lucca, Italy
[*]corresponding author(s): Christian Herff (c.herff@maastrichuniversity.nl), Laura Marras (laura.marras@imtlucca.it)

ABSTRACT

Ideally, decisions are made based on prior knowledge, which allows for informed choices. Real life, however, often requires us to make decisions arbitrarily, without sufficient information. To understand the differences between the decision-making processes in arbitrary and informed decisions, we recorded intracranial electroencephalography in a large number of cortical and subcortical areas from 5 patients during a categorization task. We demonstrate that individual decisions can be decoded from Local Field Potentials (LFPs) before motor response, in both arbitrary and informed conditions. Our analysis revealed dissimilar spatio-temporal patterns between arbitrary and informed decision-making, with arbitrary decisions being decodable in fewer brain regions and earlier in time compared to informed decisions. These results mark a crucial step towards the development of cognitive neuroprostheses and Brain-Computer Interfaces (BCIs) to support decision processes in rapid human-machine interactions, weigh decision-making confidence, and further enable neuromodulation protocols for the treatment of reward-related dysfunctions.

Introduction

Adaptive behaviors, intended as actions aimed at maximizing survival and minimizing harm, require the ability to evaluate environmental stimuli, process the possible responses, anticipate their consequences, and choose the most appropriate action[1]. Therefore, the capacity to make informed decisions is crucial to survival. Unfortunately, daily life is commonly characterized by novel situations where prior knowledge of how possible choices are mapped onto outcomes is not available. In such situations, initial reflexes and habits, followed by subsequent instrumental learning play a crucial role in creating action-outcome mappings through reward-driven exploration of the decision space[2–4]. In this trial-and-error strategy, decisions are initially arbitrary, but the information obtained through the initial decisions can later be used to guide informed decision-making.

The neural underpinnings of informed and arbitrary decision-making remain uncertain. Studies that have examined these forms of decision-making individually suggest they may recruit distinct neural circuits. In particular, informed decisions have been associated with activity in the prefrontal cortex[5], orbitofrontal cortex[6], and anterior cingulate cortex[7,8], whereas arbitrary decisions have been linked to activations in the supplementary motor area and medial frontal cortex[9–11]. However, conflicting evidence emerges from studies that have directly compared the electrophysiological responses associated with these distinct decision-making processes[12]. Some studies found that the Readiness Potential (RP) - a negative electrical component associated with motor decisions[13] - is smaller or absent in informed compared to arbitrary decisions[14,15]. In contrast, other studies reported the opposite effect[16] or no differences at all[17].

Understanding the neurophysiological signatures of informed and arbitrary decisions represents the first step toward the development of neural prosthetics such as targeted neuromodulation protocols[18] and decision-making Brain-Computer Interfaces (BCIs)[19–21]. Specifically, closed-loop neuromodulation techniques have been revealed to be promising in the treatment of disorders characterized by deficits in reward-related behaviors and decision-making, such as addiction[22–24], pathological gambling[24,25], and affective disorders[26]. However, these approaches require the identification of reliable neural features of decision-making that can be robustly decoded and are suitable to serve as control signals for stimulation. Additionally, being able to decode decisions early, before motor planning onset, might contribute to the improvement of BCIs for motor control and communication, allowing shorter reaction times and better error processing[19].

Previous studies in Non-Human Primates have shown that decision-related features can be decoded from electrophysiological recordings. Rich and Wallis were able to decode option values during binary decisions, based on single-unit activity (SUA) and local field potential (LFP) recordings from the monkey orbitofrontal cortex[27]. In a recent study, a linear decoder was applied to neural population activity from the primary motor and dorsal premotor cortices to track the deliberation process and predict the animal's choices at the single trial level[28]. These studies demonstrated the feasibility of single-trial characterization of neural features in decision-making based on small- and large-scale electric signals in animals.

In humans, early attempts at decoding decision-related variables were carried out in functional magnetic resonance imaging (fMRI) studies. Specifically, by applying multivariate techniques to fMRI recordings, multiple studies achieved successful prediction of various behavioral aspects of decision-making, including reward-related decisions[29], belief or disbelief decisions[30], action intentions[31], and single-choice outcomes[32]. However, while fMRI allows the identification of spatial activation patterns underlying decision encoding, it does not reveal insights into the temporal dynamics or the role of oscillatory activity. Furthermore, the costs, lack of portability, and low temporal resolution render fMRI unfeasible for online BCIs or treatment solutions.

To this end, a few studies have explored decoding single decisions from electroencephalography (EEG) activity in humans. In a study by Si and colleagues, the authors developed a computational framework to extract features from single-trial EEG activity and use them to predict individual decision responses in the ultimatum game[33]. They demonstrated that different decision responses could be decoded from EEG network activity on a trial-by-trial basis, supporting the idea that acceptance and rejection decisions engage different brain networks[34]. Similarly, a recent study implemented a novel feature extraction method to discriminate cooperation and aggression decisions based on EEG. The authors showed that this method can accurately classify different decision responses and that the primary source of discriminability comes from the EEG information within the frequency range below 40 Hz[35]. These findings support the role of specific rhythmic activity in decision encoding and suggest that neural fingerprints of this process can be extracted from electrophysiological activity. However, scalp EEG suffers from multiple limitations related to spatial resolution, signal-to-noise ratio, sensitivity to artifacts, and, most importantly, access to higher frequency bands[36–38], which substantially restrain its use as a potential signal for BCIs and neuroprosthetics.

Conversely, stereotactic EEG (sEEG) provides access to high-frequency LFP activity and allows to characterize the temporal patterns of cognitive processing within spatially well-defined brain sites[36,38]. In particular, high-frequency LFP power has been indicated as an optimal signal for BCI control[39,40], as it reflects integrated spiking of multiple neurons near the recording site[37,41,42] and has been shown to have comparable decoding accuracy and informative content with single-unit spikes[40,43]. Additionally, compared to SUA, using LFPs allows for up-scaling to more complex decision spaces whilst remaining in a timescale useful for BCIs, requires less computational power, and a lower data sampling rate[40]. Thiery and colleagues recorded sEEG while patients performed an oculomotor decision task involving free and instructed eye movements[44]. They showed that LFPs could be used to accurately classify saccades as instructed or free on a single-trial basis. Notably, while the brain areas engaged by free-choice and instructed actions were mostly shared, the two conditions differed in the temporal dynamics of high-frequency activity, which was found to be the most informative feature for the classification task.

Taken together, human and animal studies provide converging evidence that neural signatures of decision-making can be identified in broadband electrophysiological activity, which is best suited to extract reliable features for BCI control and neuroprosthetics applications.

In this study, we probed neurophysiological fingerprints of decision-making, focusing for the first time on informed and arbitrary decisions. Stereotactic EEG was recorded from a total of 590 contacts in 5 patients suffering from pharmaco-resistant epilepsy, during a decision-making paradigm including arbitrary and informed choices. By applying decoding analysis to broadband LFP activity, we show that the outcome of single categorization choices can be decoded with high accuracy for both arbitrary and informed decisions. Furthermore, the resolution of intracranial recordings allowed us to examine the spatial and temporal dynamics of decision decoding and identify previously unknown differences between informed and arbitrary decision-making.

Results

Participants performed a simple feedback-guided categorization task in which they had to assign each presented stimulus to one of two abstract categories (i.e., "Winning" or "Losing") using a trial-and-error strategy. Each stimulus was presented three times in a pseudo-randomized order (Fig.1a). Following the stimulus, participants were given 2 seconds to indicate the selected category by pressing the corresponding button (left or right arrow). Participants received feedback after each decision indicating whether they selected the correct category (Fig.1b). Since stimulus-category associations were random and meaningless, no task-relevant prior knowledge was available to the participants when performing the first trial of each stimulus. Therefore, whenever a new stimulus was presented, they had to simply guess (arbitrary decisions). Conversely, when stimuli were presented for the second and third time, participants were instructed to adjust their decisions based on the previously received feedback (informed decisions). Accordingly, we expected participants' performance to be at chance level during arbitrary decisions and then improve during informed decisions.

Participants shift from arbitrary to informed decisions in a categorization task

First, we tested whether feedback information was processed and integrated into the decision-making process. As expected, average accuracy across participants was at chance level (0.43) during arbitrary decisions. Performance significantly improved in the second (0.80, $p < 0.01$, Tukey's HSD test) and third (0.76, $p < 0.01$, Tukey's HSD test) compared to the first stimulus presentation (Fig.1c). These results show that participants successfully learned the correct stimulus-category association after feedback presentation and used the provided feedback information to improve their decisions. Average reaction time across participants was 335ms ±228ms after cue onset (428ms ±279ms in 1st presentation, 244ms ±188ms in 2nd presentation, and 333ms ±229ms in 3rd presentation).

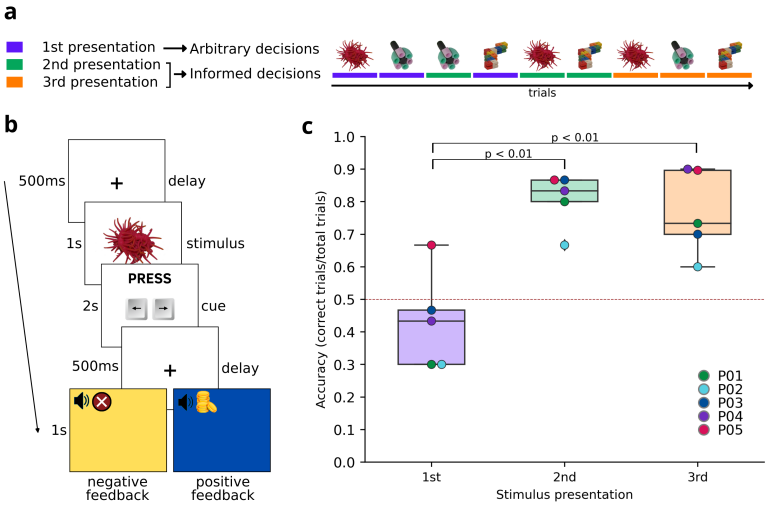


Figure 1. Feedback-guided categorization task. a, ***Schematic of an example session.*** In each session, three visual stimuli are presented three times, in randomized order. Color indicates first (purple), second (green), and third (orange) stimulus presentations. Arbitrary decision condition consists of trials in which stimuli were presented for the first time and participants were instructed to guess their category. Informed decision condition includes trials with second and third presentations of stimuli, where participants had to adjust or confirm their choices based on feedback received in previous trials. **b**, ***Schematic of example trial sequence.*** After 500ms of fixation period, a stimulus is presented for one second. Then a cue indicating to press the left (to choose "Winning" category) or right (to choose "Losing" category) arrow key appears for maximum 2s. As soon as a response is provided, a fixation cross is displayed for 500ms. Positive (blue screen and coins dropping sound) or negative (yellow screen and buzzing sound) feedback is delivered depending on whether participants chose the correct stimulus-category or not, presented for 1s. **c**, ***Task performance per participant and stimulus presentation.*** Boxplots represent performance distribution for each stimulus presentation, each dot represents one participant's average performance. The red dotted line indicates chance level.

Decoding arbitrary and informed decisions from LFP activity

Secondly, we explored whether individual upcoming decisions could be decoded from single-channel LFP activity. We trained a linear discriminant analysis (LDA) classifier on multi-band LFP power averaged over the time period of stimulus presentation (Methods). We tested decoding of decision (selected category) for each recording channel of each participant and decision condition (arbitrary or informed). Permutation test revealed that out of a total of 590 channels, the classifier was able to decode individual choices with significant accuracy (ROC AUC, $p < 0.05$, permutation tests) in 49 channels during informed decisions and 28 channels in the arbitrary decision condition (Fig.2a). Among these, performances varied across participants and channels, with ROC AUC ranging from 0.59 to 0.77. Notably, significantly accurate decision decoding was observed in at least two channels in every single participant, in both arbitrary and informed conditions. These results show that decoding decisions from multi-band LFP activity is feasible at the single channel level. This means that an upcoming decision can be predicted solely based on neural activity during the stimulus presentation.

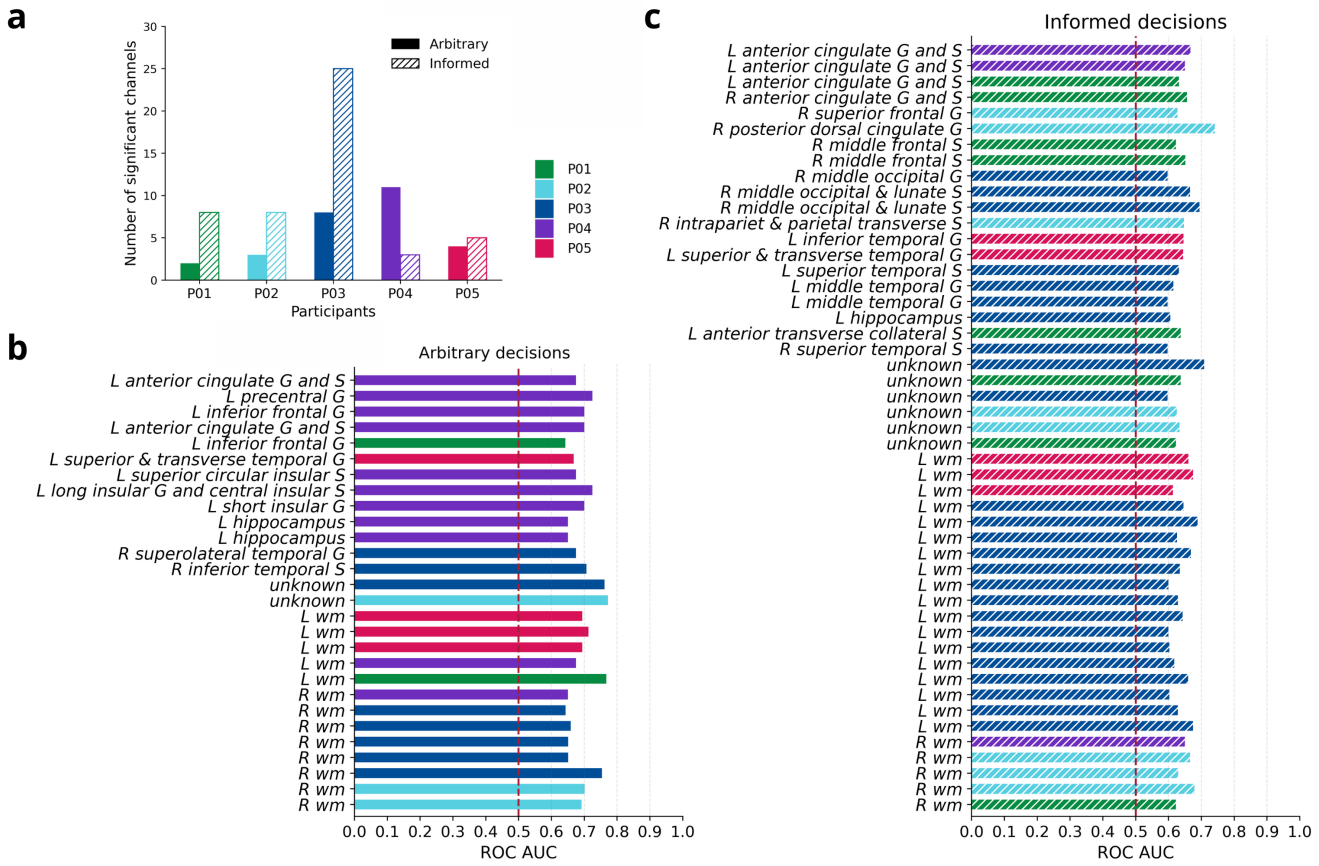


Figure 2. Decoding single-choice from single-channel multi-band LFP activity. An LDA classifier was applied to the multi-band LFP power (average over the entire stimulus period) of single channels to decode single decisions ("L" vs "W" presses), for each participant and condition separately. **a**, ***Number of informative channels per participant and condition.*** For each participant (color coded) the number of channels that resulted in significant decoding ROC AUC ($p < 0.05$, permutation tests) is shown for arbitrary (full bars) and informed (dashed bars) conditions. **b**, ***Decoding results of informative channels during arbitrary decisions.*** The bar plot shows ROC AUC scores and anatomical location of all channels that resulted in significant decoding results ($p < 0.05$, permutation tests) during arbitrary decisions. Bar colors represent participants as in **a**. L=left, R=right, G=gyrus, S=sulcus, wm=white matter. **c**, Same as in **b**. for the informed decision condition.

Spatial dynamics of decision decoding

We then looked at the anatomical location of the contacts with significant decoding accuracy and compared the results across the two decision conditions. Across all participants, only 5 channels showed significant decoding performance in both conditions (Fig.2b, Fig.3). Of these channels, two were located in the left anterior cingulate cortex, one in the left superior temporal gyrus and one in white matter. The 23 contacts that showed significant activity only during arbitrary decision-making were located

in various portions of temporal and frontal cortices, in the hippocampus and in white matter (Fig.2b, Fig.3). Finally, the 44 channels that reached significant decoding performance exclusively for informed decisions were spread out across all four cortical lobes and various subcortical structures, including occipital areas, the hippocampus and white matter (Fig.2b, Fig.3). This analysis revealed dissimilar spatial patterns with only a small overlap between informed and arbitrary decision-making.

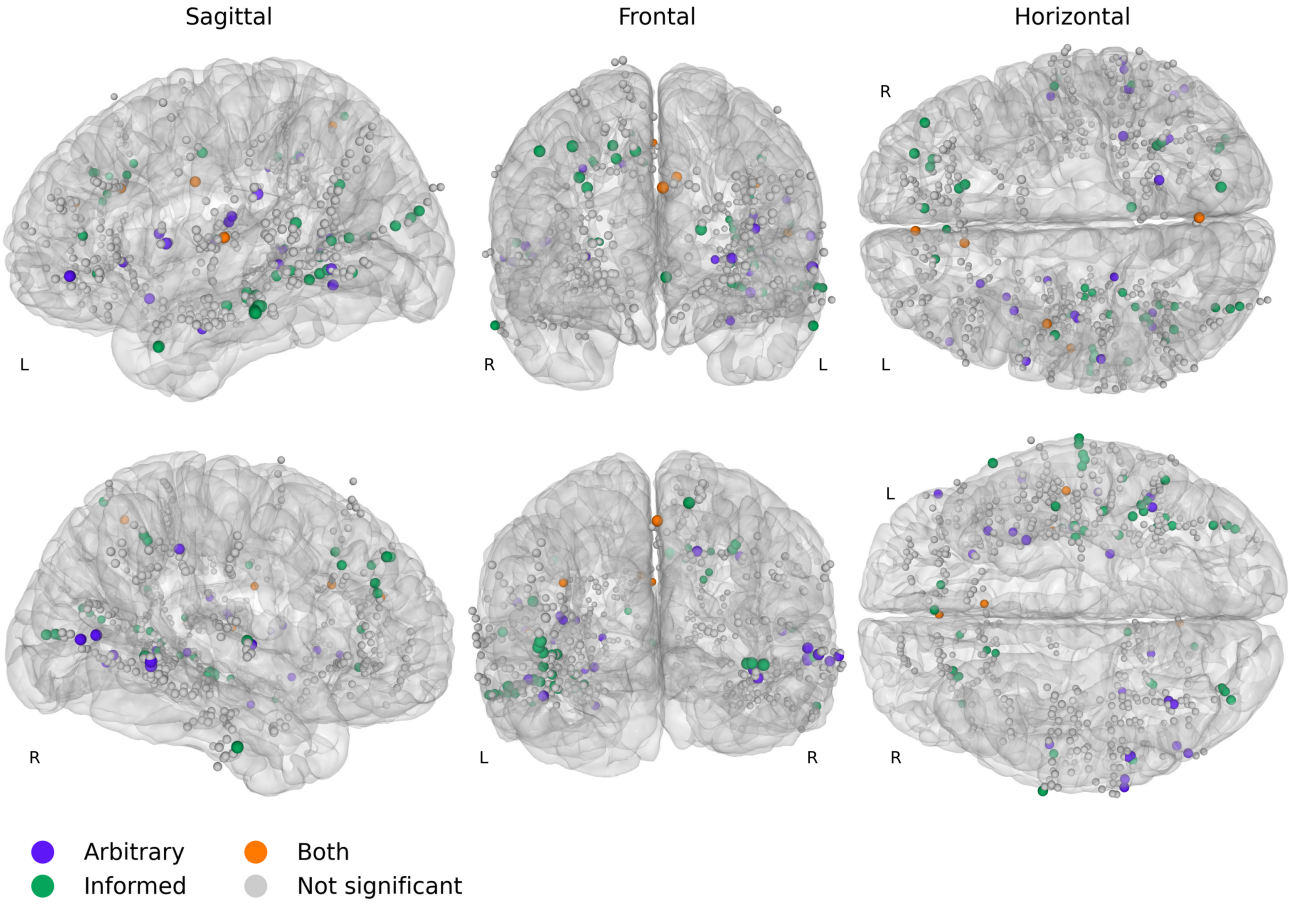


Figure 3. Spatial layout of informed and arbitrary decision decoding. Contacts anatomical location of all participants are shown in the MNI template space (obtained through warping from each individual's native MRI space) from sagittal, horizontal and coronal views. Color indicates whether the channel resulted in significant decoding ROC AUCs ($p < 0.05$, permutation tests) during arbitrary condition only (purple), informed condition only (green), informed and arbitrary conditions (orange) or none (grey).

Time course of decision decoding

Lastly, we examined the temporal dynamics of decision decoding. To achieve this, we first split the 2s time window including inter-trial delay, stimulus, and cue periods, into 39 partially overlapping windows of 100ms. Then, for each window, an LDA classifier was applied to the average multi-band LFP power (Methods). Only the channels that had significant decoding performance in the previous analysis were included in this analysis. Permutation tests showed that a total of 48 (for informed decisions) and 27 (arbitrary decisions) channels reached significant decoding accuracy (ROC AUC, $p < 0.05$, permutation tests) in at least 1 window. For further analysis, we selected the channels in which decoding accuracy was significant in at least 2 consecutive temporal windows during stimulus presentation. This criterion was met in 31 channels in classification of informed decisions and 12 channels in classification of arbitrary decisions.

As shown in Fig.4a and e, the time course of decoding accuracy varied substantially across channels. A linear regression analysis (Methods) revealed a significant interaction effect between time and condition on ROC AUC scores ($\beta = 0.002$,

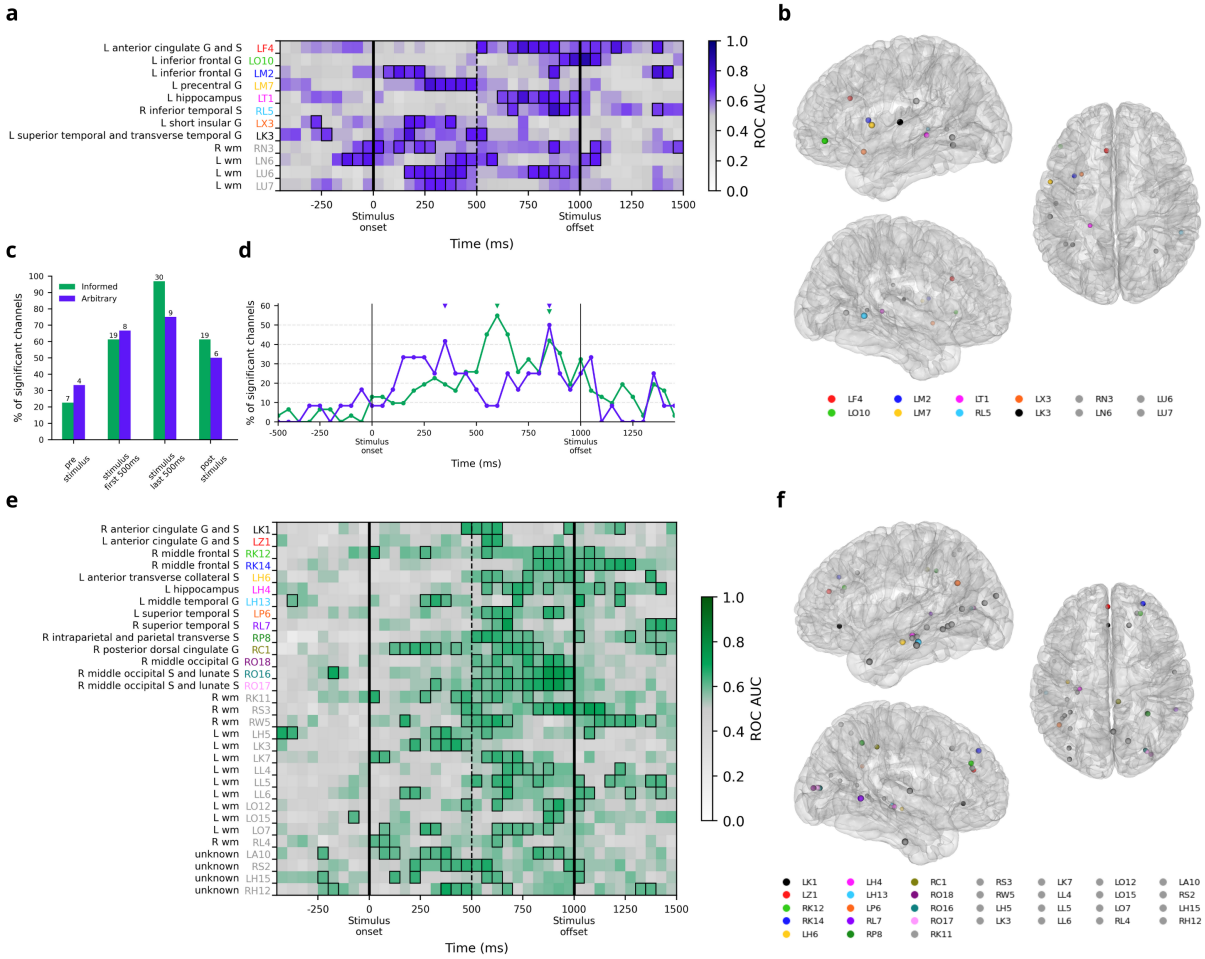


Figure 4. Spatio-temporal configuration of informed and arbitrary decision decoding. a, ***Time course of arbitrary decision decoding.*** ROC AUC scores are plotted as a function of time (time-locked to stimulus onset) for each analyzed channel. Thick borders indicate the time windows in which ROC AUC is significantly higher than chance ($p < 0.05$, permutation tests). Vertical lines signal stimulus onset and offset (full) and half-period of stimulus presentation (dashed). Colored channels' labels represent channel anatomical location in **b**; grey labels indicate channels in white matter or unknown location. **b,** ***Spatial layout of arbitrary decision decoding.*** Sagittal and horizontal view of the MNI template brain. Dots show the anatomical location of channels in **a** (color-coded). **c**, ***Percentage of significant channels across task periods.*** Percentage and absolute number (above each bar) of channels showing significant ROC AUC ($p < 0.05$, permutation tests) of informed (green) and arbitrary (purple) decisions for each task period. **d**, ***Percentage of significant channels across time.*** Time-course of the percentage of channels showing significant ROC AUC ($p < 0.05$, permutation tests) of informed (green) and arbitrary (purple) decisions. Triangles on top indicate peak locations for each condition. **e**, **f**, Same as in **a**, **b**, for informed decision condition.

$SE = 0.000$, $t = 4.73$, $p < 0.001$), suggesting that the effect of time on accuracy differs across conditions, with overall higher scores in arbitrary compared to informed decisions at earlier time points.

To assess the time effect for each condition, we first examined the percentage of significant channels in each task period, for each condition. Most channels showed significant decoding accuracy during the last 500ms of stimulus presentation ($n = 9$ in arbitrary, $n = 30$ in informed) and within the first 500ms ($n = 8$ in arbitrary, $n = 19$ in informed). Additionally, in some channels we were able to decode upcoming decision during pre-stimulus fixation, preceding motor response by more than 1 second ($n = 4$ in arbitrary, $n = 7$ in informed). Finally, some channels showed significant decoding accuracy during the post-stimulus cue period ($n = 6$ arbitrary, $n = 19$ informed) (Fig.4c). These findings support the idea that LFP activity conveys information about upcoming categorization choice while the stimulus is visually perceived and close to the corresponding

motor response, but also well before delivery of motor response.

Lastly, we focused on the percentage of significant channels at each time bin and identified the main peaks for each condition (Methods). We found two peaks for arbitrary decisions, located at 350ms and 800ms after stimulus onset, and two peaks in the informed condition, located at 600ms and 800ms after stimulus onset (Fig.4d). This analysis suggests that the time course of this representation differs between arbitrary and informed conditions, with predictive neural activity emerging earlier for arbitrary decisions compared to informed decisions.

Discussion

We showed that individual category decisions can be decoded from single-channel multi-band LFP power, with a ROC AUC of up to 0.77. This means that, at the single channel level, the average power value of different LFP frequency bands - during stimulus presentation - provides valuable information to successfully predict whether the subject will choose the "Winning" or the "Losing" category. This suggests that broadband electrophysiological activity conveys information about choice, constituting a potential mechanism by which this information is represented in the brain.

Our task allowed us to distinguish two types of decisions. In one-third of the trials, participants had to make arbitrary choices on novel stimuli where the only possible strategy was to guess, by relying on reflexes or habits, as opposed to goal-directed instrumental learning. The remaining two-thirds of the trials involved previously seen stimuli, for which participants had to retrieve the correct category based on previous choices and feedback received, thus making informed decisions. In our paradigm, the sensory input and motor output remain constant across all conditions. Thus, distinctions between the arbitrary and informed conditions are solely attributable to variations in the available decision-making information provided to the agents. We explored decoding of single choices separately for each decision condition. Importantly, we found that both arbitrary and informed types of decisions could be successfully predicted, although their representation showed distinct spatial layouts, both in terms of the number of informative channels and in terms of brain areas involved.

First of all, arbitrary choices could be decoded from activity in a smaller number of channels ($n = 28$) compared to informed decisions ($n = 49$). Interestingly, only 5 channels were informative in both conditions.

Secondly, our results enabled the distinction of three separate sets of brain regions based on which type of decisions were represented by the activity in each channel. The first set includes brain areas involved exclusively in the encoding of informed decisions, a wide range of brain areas spanning all cortical lobes, including the posterior cingulate cortex, superior and middle frontal gyri, middle occipital gyrus, intraparietal sulcus, inferior and middle temporal gyri, superior temporal sulcus, and collateral sulcus. The second set, consisting of areas that represent only arbitrary decisions, includes the precentral gyrus, inferior frontal gyrus, inferior temporal sulcus, and insula. Lastly, the third set comprises the few regions involved in the representation of both conditions: the anterior cingulate cortex, the superior temporal gyrus, and the hippocampus. Hence, it appears that each decision type is encoded by a specific ensemble of brain areas, while a limited number of areas are involved in the representation of both types. This is in line with previous studies that suggest that arbitrary and informed decision-making recruits distinct neural systems[12,14–16].

Interestingly, we also found above-chance decoding of both types of decisions in white matter contacts, in agreement with previous results showing that white matter activity contains meaningful information for decoding tasks[45].

In order to gain a deeper understanding of the temporal dynamics of decision representation, we examined the time course of decision decoding across the pre-stimulus fixation, stimulus presentation, and post-stimulus cue periods. Generally, the majority of channels exhibited the emergence of choice representation primarily during the final 500ms of stimulus presentation. Progressively fewer channels provided significant decoding in the initial 500ms of stimulus presentation, the post-stimulus cue, and the pre-stimulus fixation periods.

Importantly, we found that arbitrary and informed decisions exhibit different time courses of predictive power. While the number of channels with significant decoding accuracy first peaked at 600ms in informed decisions, for arbitrary decisions we identified an earlier first peak, located at 350ms. This observation aligns with a framework implying potential distinct mechanisms involved in the two processes[2]. Specifically, the early surge in predictive capability during arbitrary decisions may reflect the representation of reflexes or habits, whereas the later rise in informed decisions would reflect the recruitment of information from instrumental, goal-directed learning.

Both conditions showed a second peak located at 800ms. Given the average timing of motor responses (335ms $\pm$228ms following stimulus offset), it is conceivable that neural activity observed during the latter windows of stimulus presentation and post-stimulus cue periods may reflect the preparation of motor acts. Potentially, channels exhibiting significant prediction accuracy during this timeframe may convey information related to motor planning rather than abstract stimulus categorization. Nevertheless, the motor acts employed in our paradigm shared the same effector (dominant hand) and were kinetically comparable, making it challenging to distinguish from sEEG activity.

Interestingly, some channels showed significant prediction accuracy during the pre-stimulus fixation period. The predictive value of neural information for responses, even before the stimulus was presented (Fig.4), is of particular interest, as it is likely

associated with response biases and stimulus prediction. This could either be indicative of participants tracking the absolute occurrence of each stimulus, deploying statistical learning to detect potential underlying patterns of the loss-win distribution that would allow them to predict the next stimulus. Or it might be associated with a response bias.

In summary, our study demonstrated that the representation of categorization choice was observed prior to the motor response in both arbitrary and informed decisions. However, there were differences in the timing and spatial distribution of this representation between the two conditions, pointing out the involvement of distinct underlying mechanisms. These findings suggest that various aspects of decision-making, ranging from stimulus processing to action planning, are mediated by distinct mechanisms and brain regions.

To conclude, we show that both arbitrary and informed decision-making can be decoded from single-trial LFPs in sEEG recordings, which presents an important step towards neuroprostheses for decision processes and cognitive BCIs[46]. The brain-wide distributed recordings allow us to provide detailed information about the neural substrate of informed and arbitrary decision processes. Additionally, we bring robust evidence in support of the idea that single choices are represented by the brain through oscillations and shed further light on the spatial and temporal dynamics of arbitrary and informed decision-making.

Methods

Participants

Five patients (Tab.1) suffering from pharmaco-resistant epilepsy underwent implantation of sEEG electrodes for seizure source identification at the Maastricht University Medical Center. After surgery, the patients were transferred to the Epilepsy Center Kempenhaeghe to undergo the monitoring process. Here, patients agreed to voluntarily participate in our experiment. The study was approved by the Medical Ethics Review Committee of Maastricht University Medical Center and Epilepsy Center Kempenhaeghe and all patients provided written, informed consent for participating in the study.

Participant	Sex	Age	Shafts	Left Contacts	Right Contacts	Total contacts
P01	F	52	12	61	66	127
P02	M	36	13	0	122	122
P03	M	30	12	98	59	157
P04	M	36	15	109	18	127
P05	M	21	6	57	0	57

Table 1. Sex, Age, and electrode contacts information for each participant. Note that the number of contacts shown here refers to the contacts from which sEEG activity was recorded and which were included in the analyses, not the total number of implanted contacts.

Electrode implants

Patients were implanted with platinum-iridium DIXI MicroDeep electrodes (DIXI Medical, France) with a diameter of 0.8 mm, a contact length of 2 mm, and an intercontact distance of 1.5 mm. Electrode placement was decided solely based on clinical needs of each patient. T1-weighted magnetic resonance imaging (MRI) scans were acquired pre-surgery for implantation planning. Electrode locations were confirmed post-surgery through co-registration of the pre-surgery MRI with a postoperative Computerized Tomography (CT) scan. A total of 127, 122, 157, 127, and 57 contacts were recorded for each participant (Tab.1, Fig.5). Anatomical labelling was conducted through cortical parcellation based on the Destrieux atlas[47] using img-pipe[48] and FreeSurfer[49]. The electrodes labelled as "Unknown" are usually located in areas that are not identifiable through FreeSurfer parcellation, such as electrodes positioned slightly outside the cortex, or at the border between gray and white matter. For visualization purposes only, contact locations in each individual's native MRI space were warped to the average MNI space[50]. Note that for this reason, electrode positions in the MNI space might appear different from their respective labels.

Recording setup

Neural time series data were recorded with two or three stacked Micromed SD LTM amplifier (Micromed S.p.A., Italy) at 1024Hz sampling rate and referenced to a common white matter electrode. Neural data were synchronized to experimental events with LabStreamingLayer[51].

Task and stimuli

We designed a simple feedback-guided categorization task in which the goal is to assign each element of a set of visual stimuli to one of two abstract categories, adopting a trial-and-error strategy and minimizing error rates.

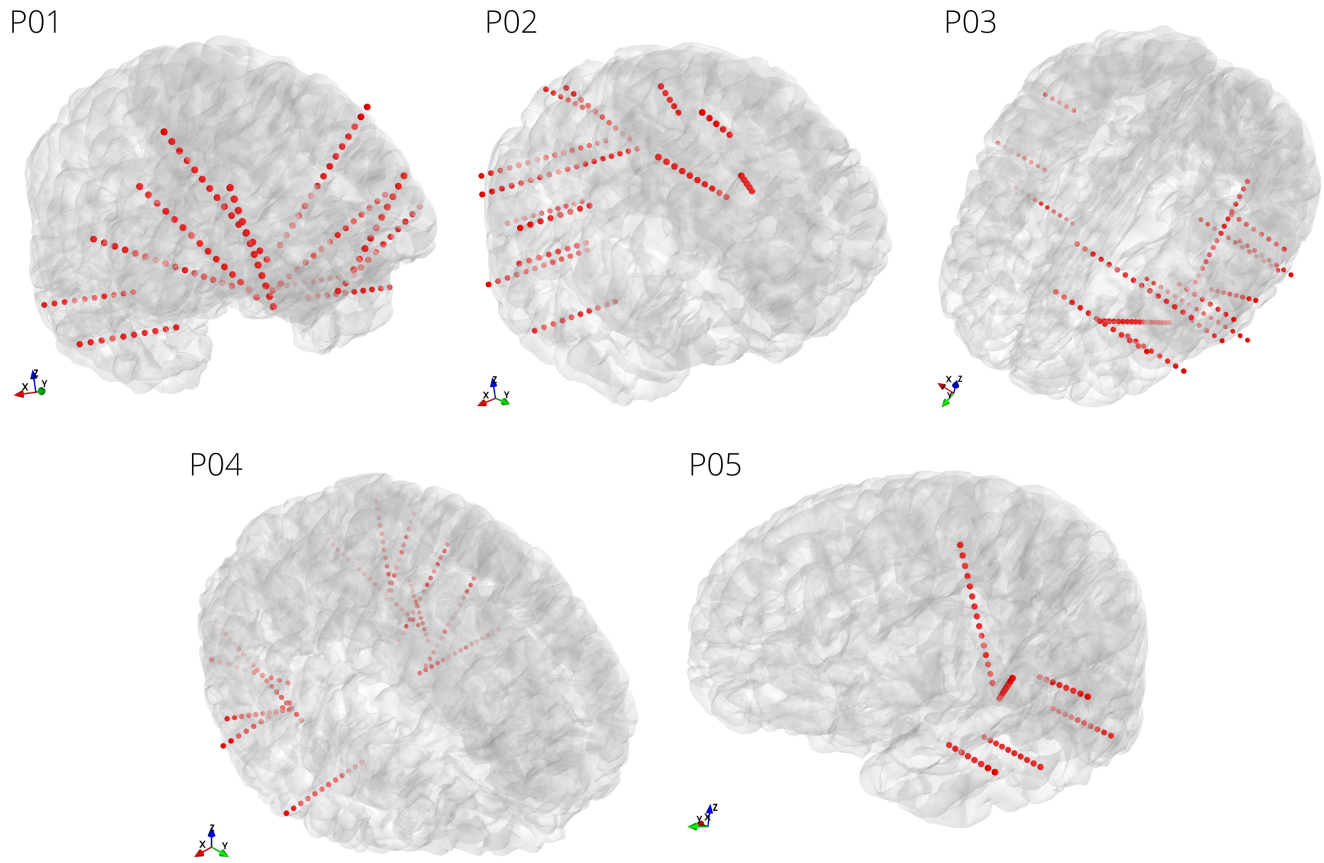


Figure 5. Electrode placement for all participants. For each participant, implanted electrodes are shown in the native MRI space. Electrode locations are determined through co-registration of pre-surgical MRI and post-surgical CT scans. Note that, based on clinical needs, data is recorded from only a subgroup of the implanted electrodes.

Stimuli were randomly chosen among the Novel Object and Unusual Name (NOUN) Database[52]. The NOUN database contains images of non-real objects, complex and distinct entities that vary along several feature dimensions in addition to shape and color. This sort of stimulus allows the minimization of preconceived valence and intrinsic stimulus information. Stimuli were randomly assigned to one of two abstract categories, named “Winning”, referred to as “W”, and “Losing”, or “L”, in equal proportions.

Each experimental session consisted of 10 blocks of 9 trials, for a total of 90 trials per session. In each block, three stimuli were presented, each for three times, in a pseudo-randomized order (Fig.1).

Each trial began with a fixation cross presented centrally on a black computer screen. After 500ms, one stimulus was presented for 1000ms. Participants were asked to pay attention to the stimulus and to withhold any response until it disappeared. Then, a cue was presented for a maximum of 2000ms, during which participants were instructed to indicate the category associated with the stimulus, by pressing with their dominant hand the right (for category “L”) or left (for category “W”) arrow key on the keyboard to indicate the category they thought was associated with this stimulus. The cue disappeared when the button was pressed. Button presses before stimulus onset were not considered. If the participant did not choose in time, a warning to speed up was shown, and the trial was considered an error. At cue offset, the fixation cross reappeared for 500ms, followed by feedback presented for 1000ms. The feedback consisted of a colored screen, either blue or yellow, indicating whether the choice was correct or incorrect (i.e., whether the participants assigned the stimulus to its correct category). Additionally, the sound of coins dropping (for correct responses) or a buzzing sound (for incorrect ones) was presented. Right after the feedback presentation, the subsequent trial started. Each trial lasted between 3000ms and 5000ms, for a total

experiment duration between 4.5 minutes and 7.5 minutes, depending on the participant's speed. Stimulus-category associations were stochastic and unknown to the participant. Thus, participants were asked to first guess and then learn the correct categories based on the received feedback.

A training session with a different set of stimuli was performed for each participant to familiarize them with the task and ensure the instructions were clear and understood.

Data preprocessing and features extraction

As a first step, sEEG signals were re-referenced using an electrode shaft referencing method[53] where each contact is re-referenced to the mean signal of all contacts within the same electrode shaft. Then, data was filtered into 5 frequency bands: delta (1-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), high-gamma (70-10 Hz), by applying a zero-phase, 4th order Butterworth bandpass IIR filter. High-gamma was included as it is known to correlate with ensemble spiking[41] and provides localized information as shown in a number of intracranial EEG decoding studies[36]. Subsequently, we estimated the envelope by extracting the magnitude of the analytic signal obtained through the Hilbert transform. Lastly, the obtained time series were epoched to extract time windows of interest for the subsequent analyses. First, a 1000ms window corresponding to stimulus presentation was extracted and used for the first decoding analysis. For the time course decoding analysis, the 2000ms epoch, including baseline (500ms, corresponding to fixation cross preceding stimulus presentation), stimulus (1000ms, corresponding to stimulus presentation) and response (500ms following stimulus offset) was segmented into 39 100ms windows with a 50ms frameshift.

Behavioral statistical analysis

Accuracy was measured as the number of correct trials over total trials for each stimulus presentation. The goal of this analysis was to establish whether accuracy differed between decisions when stimuli were presented for the first time (arbitrary) and decisions in the second and third stimulus presentations (informed). To determine the effect of stimulus presentation on accuracy scores, we performed a one-way repeated measures ANOVA and Tukey's honestly significant difference (HSD) test for multiple comparisons ($\alpha = 0.05$). Reaction times were analyzed considering the first arrow press after stimulus onset, therefore including also early responses (arrow press before cue onset). The number of trials with early responses was variable across participants (7, 50, 8, 5, 28 out of 90 trials, respectively, for P01, P02, P03, P04, and P05), occurring on average 250ms $\pm$93ms before stimulus offset.

Decoding analysis

The aim of this analysis was to explore whether we could predict upcoming decision ("W" or right presses vs "L" or left presses) based on multi-band LFP power. First, linear discriminant analysis (LDA) with shrinkage was applied to the power of 5 frequency bands averaged over the entire stimulus window (1000ms) with subsequent response ("W" or "L") used as target variable. We performed this analysis separately for the arbitrary decision condition (first stimulus presentation, 30 trials) and for the informed decision condition (second and third stimulus presentation, 60 trials). This resulted in a matrix of dimensionality = trials (30 or 60) X frequency bands (5), constituting the features of the decoder. Each decoder was evaluated with the area under the Receiver Operating Characteristic curve (ROC-AUC) using leave-one-out cross-validation.

This decoding analysis allowed us to determine which channels exhibited activity patterns containing information about decision. In a second step, we aimed to explore more in depth the time course of decision decoding in these channels. We repeated the previous analyses for each of the 39 time windows obtained by the segmentation of the 2000ms period including baseline, stimulus presentation and response (see Data preprocessing and feature extraction), for arbitrary and informed decisions separately. Here, features consisted of the average power over the 100ms windows, for each frequency band and trial. All decoding analyses were implemented in Python with scikit-learn package[54].

Permutation tests

To assess statistical significance of decoding performance we used non-parametric permutation tests[55]. First, for each participant and decoding analysis, the distribution under the null hypothesis that there is no dependency between features and target variable was estimated. For the first decoding analysis (average over the entire stimulus period), this was achieved by repeating the following steps 1000 times: randomly select one channel, permute the labels, estimate the decoding ROC AUC with leave-one-out cross-validation. At the end of this permutation process, we obtained a distribution of 1000 randomized sample decoding results. Differently, for the time-course decoding analyses, the null distribution was estimated separately for each channel, by randomly selecting one among the 39 time-windows, permuting the labels and performing the decoding analysis for a total of 1000 times.

Then, each model p-value was estimated by calculating the proportion of scores in the permuted data that are higher than the score observed in the original data. Only in models that got a value of $p < 0.05$ the null hypothesis was rejected, and the decoder was considered significant.

Additional analyses

To test for differences in decoding accuracy between conditions over time, we conducted a regression analysis, modeling ROC AUC as a function of Time and Condition ($R^2 = 0.055$, $F(3, 1673) = 32.51$, $p < 0.001$). The results revealed a significant main effect of condition on ROC AUC scores ($\beta = -0.05$, $SE = 0.01$, $t = -5.27$, $p < 0.001$), a non-significant main effect of time on accuracy ($\beta = 0.0001$, $SE = 0.000$, $t = 0.39$, $p = 0.70$), and a significant interaction effect between time and condition on accuracy ($\beta = 0.002$, $SE = 0.000$, $t = 4.73$, $p < 0.001$).

In order to further depict the specific differences in the time course of decoding accuracy between conditions, we conducted peak detection analysis on the percentage of significant channels over time bins. We defined peaks as the local maxima with a minimum height of 35, for each condition.

References

1. Rangel, A., Camerer, C. & Montague, P. R. Neuroeconomics: The neurobiology of value-based decision-making. *Nat. Rev. Neurosci.* **9**, 545 (2008).
2. Balleine, B. W. The meaning of behavior: discriminating reflex and volition in the brain. *Neuron* **104**, 47–62 (2019).
3. O'Doherty, J. P., Cockburn, J. & Pauli, W. M. Learning, reward, and decision making. *Annu. review psychology* **68**, 73–100 (2017).
4. Morris, R. W., Dezfouli, A., Griffiths, K. R., Le Pelley, M. E. & Balleine, B. W. The neural bases of action-outcome learning in humans. *J. Neurosci.* **42**, 3636–3647 (2022).
5. Wallis, J. D. & Miller, E. K. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* **18**, 2069–2081 (2003).
6. Wallis, J. D. *et al.* Orbitofrontal cortex and its contribution to decision-making. *Annu. review neuroscience* **30**, 31–56 (2007).
7. Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J. & Rushworth, M. F. Optimal decision making and the anterior cingulate cortex. *Nat. neuroscience* **9**, 940–947 (2006).
8. Hart, E. E., Blair, G. J., O'Dell, T. J., Blair, H. T. & Izquierdo, A. Chemogenetic modulation and single-photon calcium imaging in anterior cingulate cortex reveal a mechanism for effort-based decisions. *J. Neurosci.* **40**, 5628–5643 (2020).
9. Soon, C. S., Brass, M., Heinze, H.-J. & Haynes, J.-D. Unconscious determinants of free decisions in the human brain. *Nat. neuroscience* **11**, 543–545 (2008).
10. Fried, I., Mukamel, R. & Kreiman, G. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* **69**, 548–562 (2011).
11. Soon, C. S., He, A. H., Bode, S. & Haynes, J.-D. Predicting free choices for abstract intentions. *Proc. Natl. Acad. Sci.* **110**, 6217–6222 (2013).
12. Bold, J. l., Mudrik, L. & Maoz, U. How are arbitrary and deliberate decisions similar and different? In Maoz, U. & Sinnott-Armstrong, W. (eds.) *Free Will: Philosophers and Neuroscientists in Conversation*, 0, DOI: 10.1093/oso/9780197572153.003.0019 (Oxford University Press, 2022).
13. Schurger, A., Hu, P. B., Pak, J. & Roskies, A. L. What Is the Readiness Potential? *Trends Cogn. Sci.* **25**, 558–570, DOI: 10.1016/j.tics.2021.04.001 (2021).
14. Maoz, U., Yaffe, G., Koch, C. & Mudrik, L. Neural precursors of decisions that matter—an erp study of deliberate and arbitrary choice. *Elife* **8**, e39787 (2019).
15. Travers, E. & Haggard, P. The readiness potential reflects the internal source of action, rather than decision uncertainty. *Eur. J. Neurosci.* **53**, 1533–1544 (2021).
16. Travers, E., Friedemann, M. & Haggard, P. The readiness potential reflects planning-based expectation, not uncertainty, in the timing of action. *Cogn. neuroscience* **12**, 14–27 (2021).
17. Parés-Pujolràs, E., Travers, E., Ahmetoglu, Y. & Haggard, P. Evidence accumulation under uncertainty-a neural marker of emerging choice and urgency. *NeuroImage* **232**, 117863 (2021).
18. Orsini, C. A., Hernandez, C. M., Bizon, J. L. & Setlow, B. Deconstructing value-based decision making via temporally selective manipulation of neural activity: Insights from rodent models. *Cogn. Affect. & Behav. Neurosci.* **19**, 459–476 (2019).

19. Mirabella, G. & Lebedev, M. A. Interfacing to the brain's motor decisions. *J. Neurophysiol.* **117**, 1305–1319, DOI: 10.1152/jn.00051.2016 (2017).

20. Andersen, R. A., Hwang, E. J. & Mulliken, G. H. Cognitive neural prosthetics. *Annu. review psychology* **61**, 169 (2010).

21. Pesaran, B., Musallam, S. & Andersen, R. Cognitive neural prosthetics. *Curr. Biol.* **16**, R77–R80 (2006).

22. Habelt, B., Arvaneh, M., Bernhardt, N. & Minev, I. Biomarkers and neuromodulation techniques in substance use disorders. *Bioelectron. medicine* **6**, 1–17 (2020).

23. Luigjes, J., Segrave, R., de Joode, N., Figee, M. & Denys, D. Efficacy of invasive and non-invasive brain modulation interventions for addiction. *Neuropsychol. review* **29**, 116–138 (2019).

24. Fecteau, S., Fregni, F., Boggio, P. S., Camprodon, J. A. & Pascual-Leone, A. Neuromodulation of decision-making in the addictive brain. *Subst. use & misuse* **45**, 1766–1786 (2010).

25. Goudriaan, A. E. & Schluter, R. S. Non-invasive neuromodulation in problem gambling: What are the odds? *Curr. Addict. Reports* **6**, 165–174 (2019).

26. Bina, R. W. & Langevin, J.-P. Closed loop deep brain stimulation for ptsd, addiction, and disorders of affective facial interpretation: review and discussion of potential biomarkers and stimulation paradigms. *Front. Neurosci.* **12**, 300 (2018).

27. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nat. neuroscience* **19**, 973–980 (2016).

28. Peixoto, D. *et al.* Decoding and perturbing decision states in real time. *Nature* **591**, 604–609 (2021).

29. Hampton, A. N. & O'doherty, J. P. Decoding the neural substrates of reward-related decision making with functional mri. *Proc. national Acad. sciences* **104**, 1377–1382 (2007).

30. Douglas, P. K. *et al.* Single trial decoding of belief decision making from eeg and fmri data using independent components features. *Front. human neuroscience* **7**, 392 (2013).

31. Gallivan, J. P., McLean, D. A., Valyear, K. F., Pettypiece, C. E. & Culham, J. C. Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *J. Neurosci.* **31**, 9599–9610 (2011).

32. Haynes, J.-D. Decoding and predicting intentions. *Annals New York Acad. Sci.* **1224**, 9–21 (2011).

33. Si, Y. *et al.* Predicting individual decision-making responses based on single-trial eeg. *NeuroImage* **206**, 116333 (2020).

34. Si, Y. *et al.* Different decision-making responses occupy different brain networks for information processing: a study based on eeg and tms. *Cereb. Cortex* **29**, 4119–4129 (2019).

35. Huang, Z. *et al.* Discriminability of single-trial eeg during decision-making of cooperation or aggression: a study based on machine learning. *Med. & Biol. Eng. & Comput.* 1–11 (2022).

36. Herff, C., Krusienski, D. J. & Kubben, P. The potential of stereotactic-eeg for brain-computer interfaces: current progress and future directions. *Front. neuroscience* **14**, 123 (2020).

37. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nat. reviews neuroscience* **13**, 407–420 (2012).

38. Parvizi, J. & Kastner, S. Promises and limitations of human intracranial electroencephalography. *Nat. neuroscience* **21**, 474–483 (2018).

39. Zhuang, J., Truccolo, W., Vargas-Irwin, C. & Donoghue, J. P. Decoding 3-d reach and grasp kinematics from high-frequency local field potentials in primate primary motor cortex. *IEEE Transactions on Biomed. Eng.* **57**, 1774–1784 (2010).

40. Brincat, S. *et al.* Which neural signals are optimal for brain-computer interface control? In *Proceedings of the Fifth International Brain-Computer Interface Meeting*, 104 (2013).

41. Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J. & Hsiao, S. S. Neural correlates of high-gamma oscillations (60–200 hz) in macaque local field potentials and their potential implications in electrocorticography. *J. Neurosci.* **28**, 11526–11536 (2008).

42. Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS biology* **9**, e1000610 (2011).

43. Mehring, C. *et al.* Inference of hand movements from local field potentials in monkey motor cortex. *Nat. neuroscience* **6**, 1253–1254 (2003).

44. Thiery, T. *et al.* Decoding the neural dynamics of free choice in humans. *PLoS Biol.* **18**, e3000864 (2020).

45. Mercier, M. R. *et al.* Evaluation of cortical local field potential diffusion in stereotactic electro-encephalography recordings: A glimpse on white matter signal. *NeuroImage* **147**, 219–232, DOI: https://doi.org/10.1016/j.neuroimage.2016.08.037 (2017).

46. Wallis, J. D. Decoding cognitive processes from neural ensembles. *Trends cognitive sciences* **22**, 1091–1102 (2018).

47. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53**, 1–15, DOI: https://doi.org/10.1016/j.neuroimage.2010.06.010 (2010).

48. Hamilton, L. S., Chang, D. L., Lee, M. B. & Chang, E. F. Semi-automated anatomical labeling and inter-subject warping of high-density intracranial recording electrodes in electrocorticography. *Front. Neuroinformatics* **11**, 62 (2017).

49. Fischl, B. Freesurfer. *Neuroimage* **62**, 774–781 (2012).

50. Mazziotta, J. C. *et al.* A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* **2**, 89–101 (1995).

51. Kothe, C. Lab streaming layer (lsl). *https://github. com/sccn/labstreaminglayer* **26**, 2015 (2014).

52. Horst, J. S. & Hout, M. C. The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behav. Res. Methods* **48**, 1393–1409, DOI: 10.3758/s13428-015-0647-3 (2016).

53. Li, G. *et al.* Optimal referencing for stereo-electroencephalographic (seeg) recordings. *NeuroImage* **183**, 327–335 (2018).

54. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON* 6 (2011).

55. Ojala, M. & Garriga, G. C. Permutation tests for studying classifier performance. *J. machine learning research* **11** (2010).

Acknowledgements

This publication is part of the project DESIS (with project number VI.Veni.194.021) which was awarded by the Dutch Research Council (NWO) to C.H..

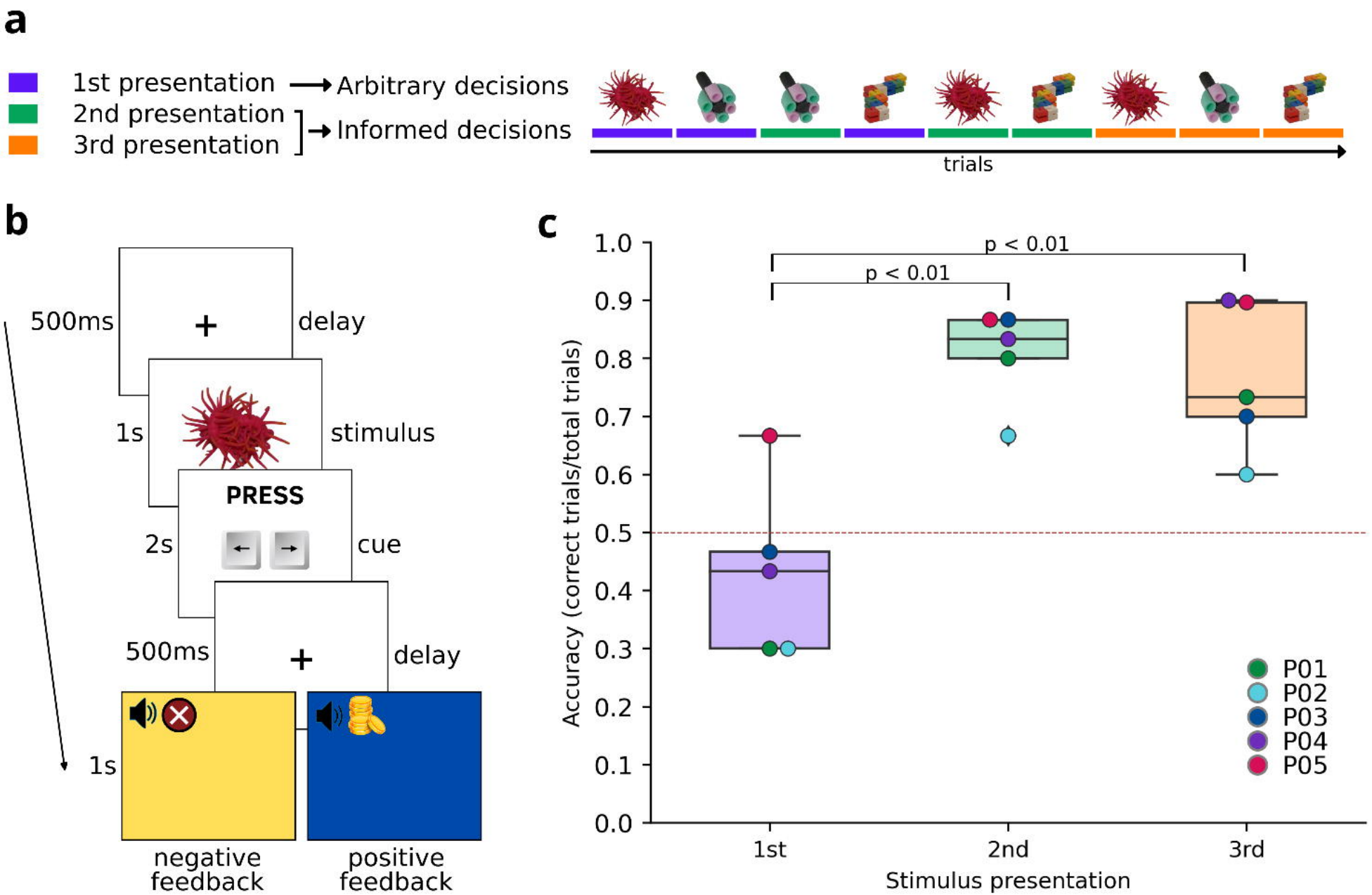
Author contributions statement

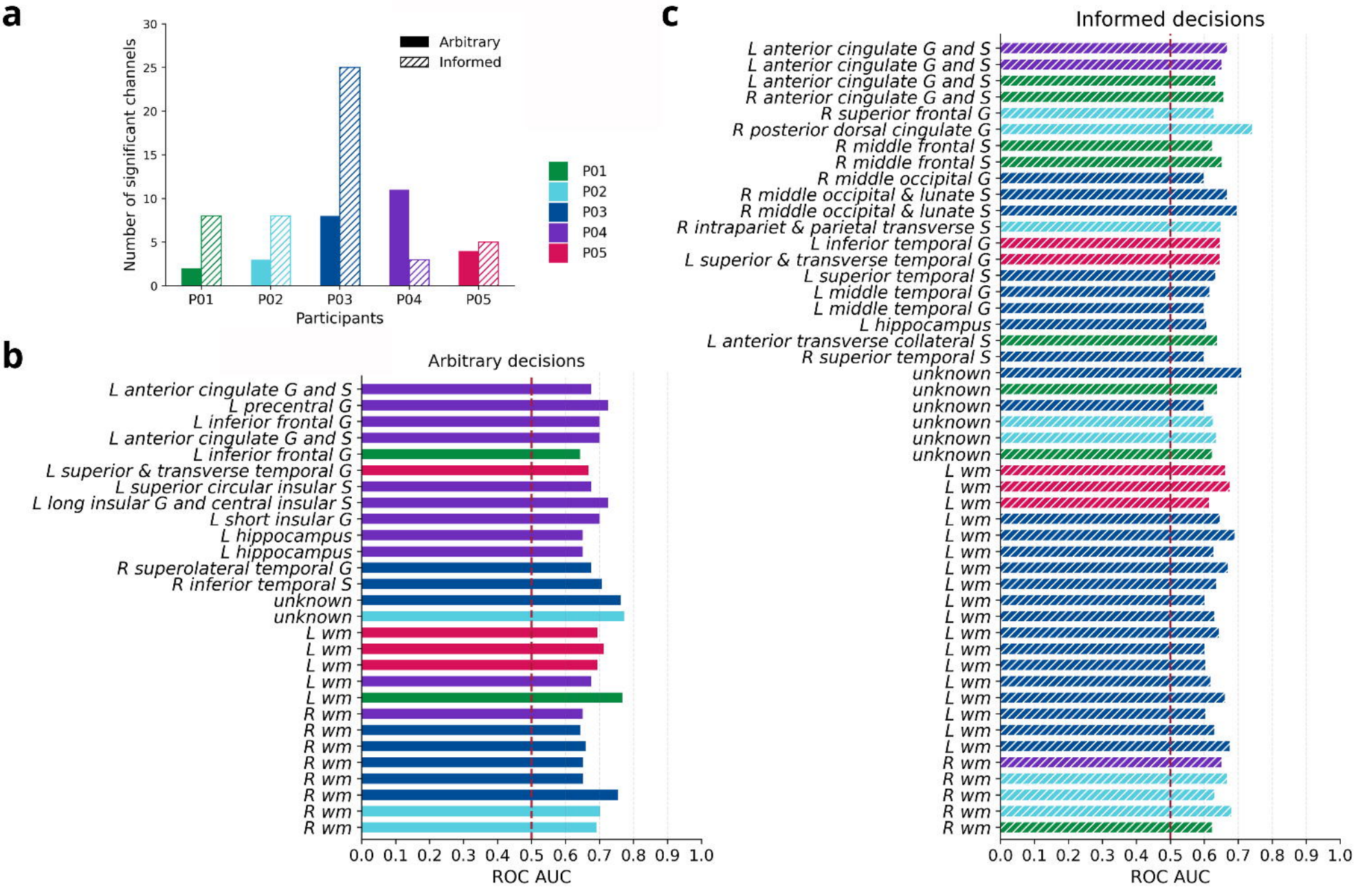
L.M., M.L.F.J., S.A.H and C.H. designed the experiment. M.V., J.P.D and M.C.O. recorded the data. S.T., L.W., A.J.C and P.K included the participants. M.V. and S.G. localized the electrodes. L.M. and C.H. analyzed the data and wrote the manuscript. All authors reviewed the manuscript.

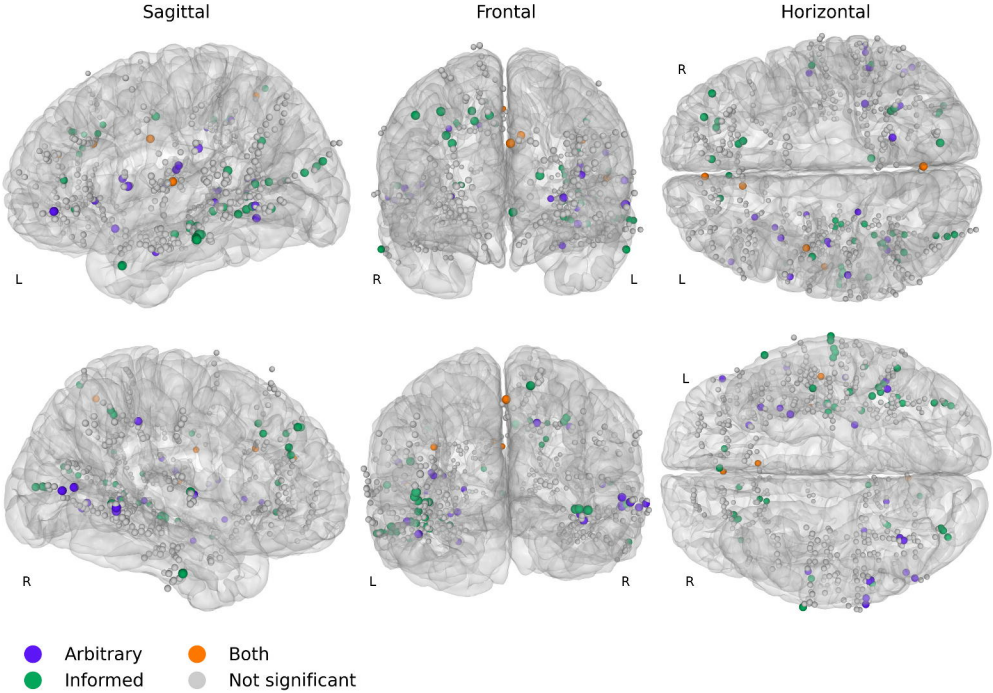
Additional information

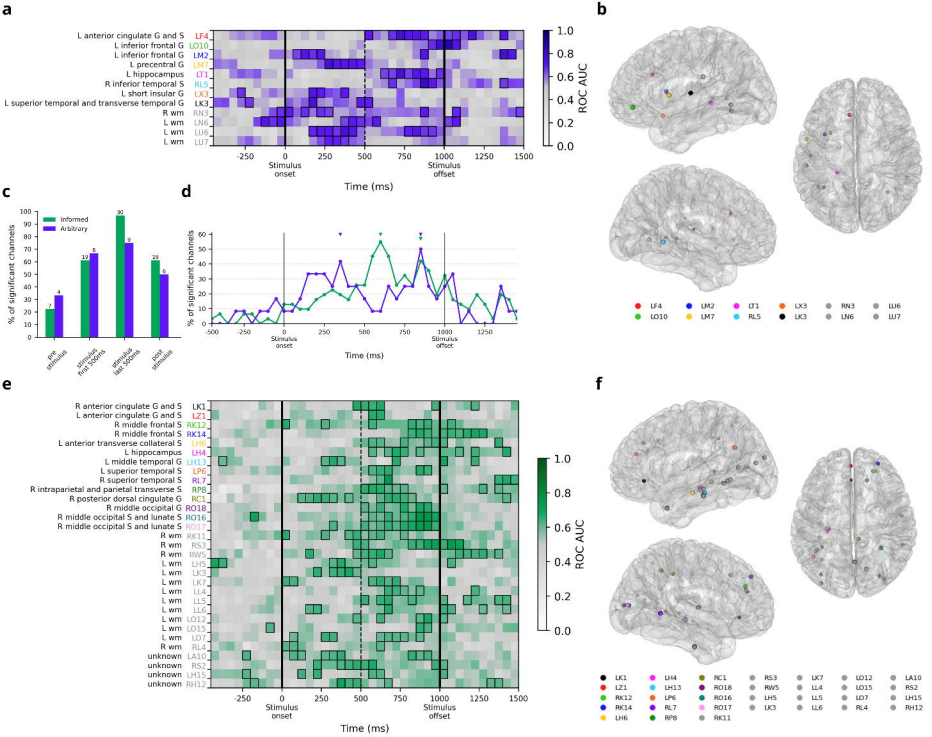
Code used in this study is available at https://github.com/Laura Marras/DecodingDecisionsSEEG. All **data** used in this study is available on https://osf.io/uaq4n/.

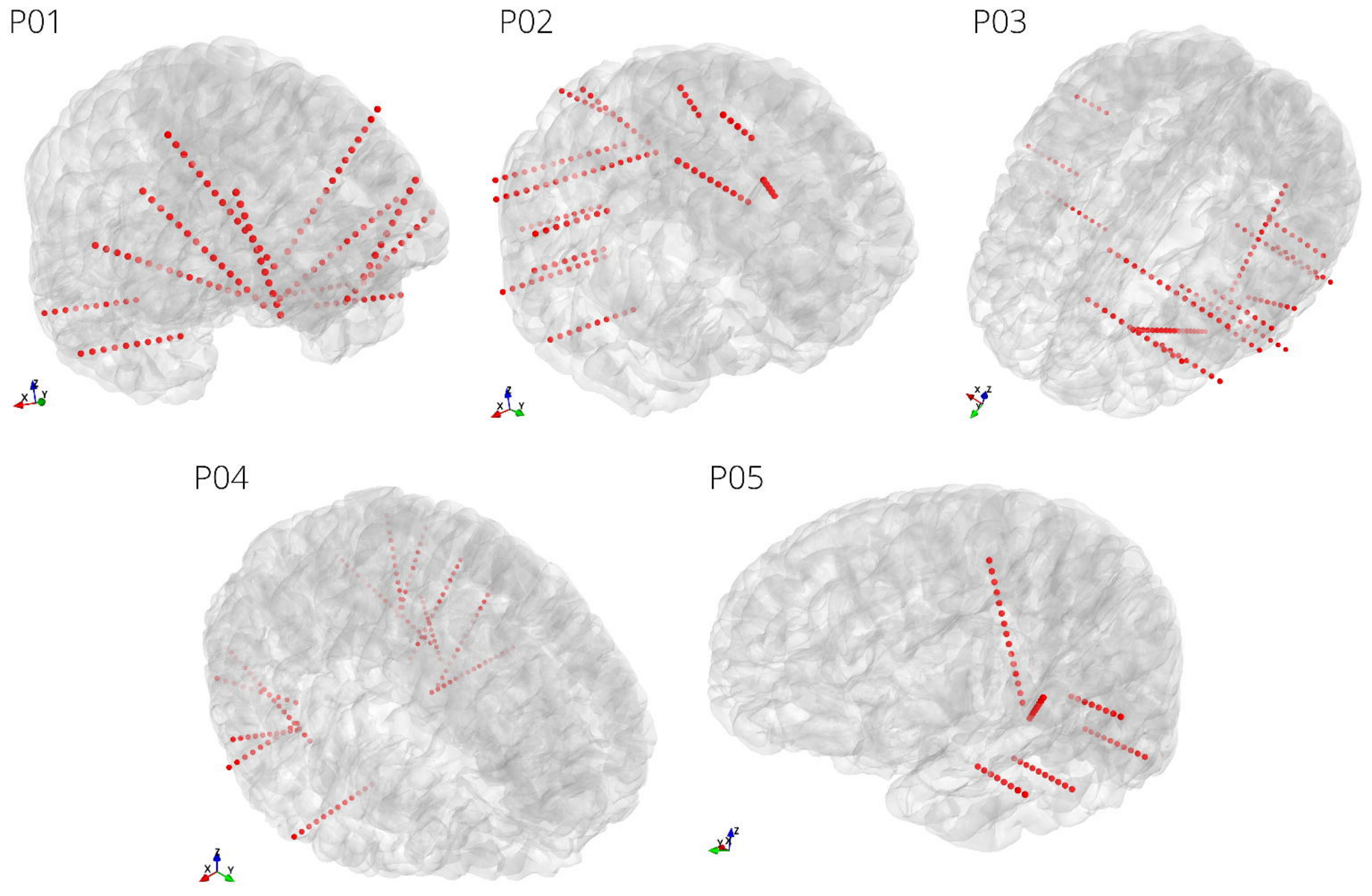
Competing interests. The authors declare no competing interests.


a
1st presentation
2nd presentation
3rd presentation
Arbitrary decisions
Informed decisions
trials
b
500ms
delay
1s
stimulus
PRESS
2s
cue
500ms
delay
1s
negative feedback
positive feedback
c
p < 0.01
p < 0.01
Accuracy (correct trials/total trials)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1st
2nd
3rd
Stimulus presentation
P01
P02
P03
P04
P05


a
Arbitrary
Informed
Number of significant channels
0
5
10
15
20
25
30
P01
P02
P03
P04
P05
Participants
P01
P02
P03
P04
P05
b
Arbitrary decisions
L anterior cingulate G and S
L precentral G
L inferior frontal G
L anterior cingulate G and S
L inferior frontal G
L superior & transverse temporal G
L superior circular insular S
L long insular G and central insular S
L short insular G
L hippocampus
L hippocampus
R superolateral temporal G
R inferior temporal S
unknown
unknown
L wm
L wm
L wm
L wm
L wm
R wm
R wm
R wm
R wm
R wm
R wm
R wm
R wm
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ROC AUC
c
Informed decisions
L anterior cingulate G and S
L anterior cingulate G and S
L anterior cingulate G and S
R anterior cingulate G and S
R superior frontal G
R posterior dorsal cingulate G
R middle frontal S
R middle frontal S
R middle occipital G
R middle occipital & lunate S
R middle occipital & lunate S
R intrapariet & parietal transverse S
L inferior temporal G
L superior & transverse temporal G
L superior temporal S
L middle temporal G
L middle temporal G
L hippocampus
L anterior transverse collateral S
R superior temporal S
unknown
unknown
unknown
unknown
unknown
unknown
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
L wm
R wm
R wm
R wm
R wm
R wm
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ROC AUC


Sagittal
Frontal
Horizontal
L
R
L
R
L
R
R
L
R
L
R
Arbitrary
Informed
Both
Not significant


a
L anterior cingulate G and S LF4
L inferior frontal G LO10
L inferior frontal G LM2
L precentral G LM7
L hippocampus LT1
R inferior temporal S RL5
L short insular G LX3
L superior temporal and transverse temporal G LK3
R wm RN3
L wm LN6
L wm LU6
L wm LU7
-250 0 250 500 750 1000 1250 1500
Stimulus onset
Stimulus offset
Time (ms)
ROC AUC
1.0 0.8 0.6 0.4 0.2 0.0
b
LF4 LO10 LM2 LM7 LT1 RL5 LX3 LK3 RN3 LN6 LU6 LU7
c
Informed
Arbitrary
% of significant channels
pre stimulus
stimulus first 500ms
stimulus last 500ms
post stimulus
d
% of significant channels
Stimulus onset
Stimulus offset
Time (ms)
e
R anterior cingulate G and S LK1
L anterior cingulate G and S LZ1
R middle frontal S RK12
R middle frontal S RK14
L anterior transverse collateral S LH6
L hippocampus LH4
L middle temporal G LH13
L superior temporal S LP6
R superior temporal S RL7
R intraparietal and parietal transverse S RP8
R posterior dorsal cingulate G RC1
R middle occipital G RO18
R middle occipital S and lunate S RO16
R middle occipital S and lunate S RO17
R wm RK11
R wm RS3
R wm RW5
L wm LH5
L wm LK3
L wm LK7
L wm LL4
L wm LL5
L wm LL6
L wm LO12
L wm LO15
L wm LO7
R wm RL4
unknown LA10
unknown RS2
unknown LH15
unknown RH12
Stimulus onset
Stimulus offset
Time (ms)
ROC AUC
f
LK1 LZ1 RK12 RK14 LH6 LH4 LH13 LP6 RL7 RP8 RC1 RO18 RO16 RO17 RK11 RS3 RW5 LH5 LK3 LK7 LL4 LL5 LL6 LO12 LO15 LO7 RL4 LA10 RS2 LH15 RH12


P01
P02
P03
P04
P05