HAL open science

Beta Oscillations in Monkey Striatum Encode Reward Prediction Error Signals

Ruggero Basanisi, Kevin Marche, Etienne Combrisson, Paul Apicella, Andrea Brovelli

To cite this version:

Ruggero Basanisi, Kevin Marche, Etienne Combrisson, Paul Apicella, Andrea Brovelli. Beta Oscillations in Monkey Striatum Encode Reward Prediction Error Signals. Journal of Neuroscience, 2023, 43 (18), pp.3339-3352. 10.1523/JNEUROSCI.0952-22.2023 . hal-04088261

HAL Id: hal-04088261

https://hal.science/hal-04088261v1

Submitted on 10 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Copyright - All rights reserved

Beta oscillations in monkey striatum encode reward prediction error signals

Basanisi, R.[1 *], Marche, K.[1, 2 *], Combrisson, E.[1], Apicella, P.[1 §], Brovelli, A.[1 §]

[1] Institut de Neurosciences de la Timone, Aix Marseille Université, UMR 7289 CNRS, 13005, Marseille, France
[2] Wellcome Center for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom.

[*] R.B. and K.M. contributed equally to this work
[§] P.A. and A.B are co-senior authors

Running title
Striatal correlates of prediction errors

Corresponding authors:
Andrea Brovelli
andrea.brovelli@univ-amu.fr
Institut de Neurosciences de la Timone (INT),
UMR 7289 CNRS, Aix Marseille University,
Campus Santé Timone,
27 Bd. Jean Moulin,
13385 Marseille, France

Ruggero Basanisi
ruggero.basanisi@gmail.com
Institut de Neurosciences de la Timone (INT),
UMR 7289 CNRS, Aix Marseille University,
Campus Santé Timone,
27 Bd. Jean Moulin,
13385 Marseille, France

Keywords
Basal ganglia, local field potentials, choice behavior, learning

Abstract (210/250 words max)

Reward prediction error (RPE) signals are crucial for reinforcement learning and decision making as they quantify the mismatch between predicted and obtained rewards. RPE signals are encoded in the neural activity of multiple brain areas, such as midbrain dopaminergic neurons, prefrontal cortex and striatum. However, it remains unclear how these signals are expressed through anatomically and functionally distinct subregions of the striatum. In the current study, we examined to which extent RPE signals are represented across different striatal regions. To do so, we recorded local field potentials (LFPs) in sensorimotor, associative, and limbic striatal territories of two male rhesus monkeys performing a free-choice probabilistic learning task. The trial-by-trial evolution of RPE during task performance was estimated using a reinforcement learning model fitted on monkeys' choice behavior. Overall, we found that changes in beta-band oscillations (15-35 Hz), after the outcome of the animal's choice, are consistent with RPE encoding. Moreover, we provide evidence that the signals related to RPE are more strongly represented in the ventral (limbic) than dorsal (sensorimotor and associative) part of the striatum. To conclude, our results suggest a relationship between striatal beta oscillations and the evaluation of outcomes based on RPE signals and highlight a major contribution of the ventral striatum to the updating of learning processes.

Significance Statement (120/120 words max)

Reward prediction error (RPE) signals are crucial for reinforcement learning and decision making as they quantify the mismatch between predicted and obtained rewards. Current models suggest that RPE signals are encoded in the neural activity of multiple brain areas, including the midbrain dopaminergic neurons, prefrontal cortex and striatum. However, it remains elusive whether RPEs recruit anatomically and functionally distinct subregions of the striatum. Our study provides evidence that RPE-related modulations in LFP power are dominant in the striatum. In particular, they are stronger in the rostro-ventral rather than the caudo-dorsal striatum. Our findings contribute to a better understanding of the role of striatal territories in reward-based learning and may be relevant for neuropsychiatric and neurological diseases that affect striatal circuits.

Acknowledgements

We thank L. Renaud for assistance with monkey surgery and Dr. M. Esclapez for help with histology. RB, EC and AB received funding from the *Agence National de la Recherche* (ANR-18-CE28-0016). RB acknowledges support through a PhD Scholarship awarded by the Neuroschool. EC 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). PA and KM received funding from the *Agence Nationale de la Recherche* (ANR-11-BSV4-006). Funding for KM was partially provided by Association Française du Syndrome de Gilles de la Tourette. The *Centre de Calcul Intensif* of the Aix-Marseille University is acknowledged for granting access to its high performance computing resources.

1. Introduction (552/650 words max)

The striatum is the major component of the basal ganglia, and it plays a key role in reward-guided learning under the influence of ascending dopaminergic projections from the ventral midbrain. Indeed, dopaminergic neurons are known to encode the difference between received and expected rewards, the so-called reward prediction error (RPE) (Schultz, 2007; Fujiyama et al., 2015; Schultz, 2016a, 2016b), which is crucial for updating action values in reinforcement learning models (Sutton and Barto, 1998). Previous neurophysiological studies on primates' and rodents' striatum have shown that subsets of output neurons (Roesch et al., 2009; Oyama et al., 2010; Asaad and Eskandar, 2011) and putative interneurons (Apicella et al., 2009; Stalnaker et al., 2012) may carry RPE signals to promote reward-guided learning. Functional neuroimaging studies in humans have also highlighted the role of the striatum in encoding RPEs (O'Doherty, 2004, 2007; Bray and O'Doherty, 2007; Brovelli et al., 2008; Valentin and O'Doherty, 2009; Park et al., 2012; Kumar et al., 2018; Pine et al., 2018; Calderon et al., 2021) with a prominent contribution of the ventral striatum, including the nucleus accumbens (O'Doherty, 2004; Abler et al., 2006; O'Doherty, 2007; Hare et al., 2008). Given the functional specialization of striatal regions based on the segregation of afferent input from cortical and limbic regions (Parent and Hazrati, 1995; Haber, 2003), an important question is whether the processing of RPE signal displays any degree of anatomical specificity and a functional gradient along the sensorimotor to limbic axis.

Among the measures of neural activity that may serve as physiological markers for RPEs in different subdivisions of the striatum, local field potential (LFPs) are a good candidate, because they reflect synchronous changes in activity of neuronal populations at a finer time-scale and with a greater anatomical resolution than functional neuroimaging techniques (Goldberg, 2004; Brown and Williams, 2005; Buzsáki, 2006). A large body of evidence from animal electrophysiology has shown that LFP oscillations can be recorded from the striatum. In particular, striatal oscillatory activity in the beta-band (typically about 15–30 Hz) has been linked to task performance, including motor and nonmotor aspects of behavior in both rodents (Berke et al., 2004; Leventhal et al., 2012; Schmidt et al., 2013) and monkeys (Courtemanche et al., 2003; Bartolo et al., 2014). In addition to movement control, striatal beta-band modulation has been associated with motivational and cognitive processes, such as reinforcement learning (Feingold et al., 2015), attention (Banaie Boroujeni et al., 2020), cues utilization for action selection (Leventhal et al., 2012) reward expectation and detection (Howe et al., 2011), including reward valuation (Schwerdt et al., 2020). Moreover, some studies have pointed out that striatal beta oscillations and their relation to motor and reward processing may occur in a regionally-dependent manner (Howe et al., 2011; Schwerdt et al., 2020). Nevertheless, it remains unclear whether RPE signals during the processing of action outcomes may influence striatal beta activity.

In the present study, we recorded LFPs from different sites across the striatum of two macaque monkeys trained on a free-choice probabilistic learning task. Using a behavioral-modeling approach for the analysis of monkeys' choice behavior, we found that LFP's beta-band oscillations are related to RPE. The results show that beta-band correlates of RPE signals are differently modulated along an axis defined from the rostro-ventral to the caudo-dorsal striatum, suggesting a dominant RPE component in the first, rather than the latter part.

2. Material and Methods

2.1. Experimental procedure and data acquisition

2.1.1. Experimental setup and behavioral data

Two male adult rhesus monkeys (*Macaca mulatta*), monkeys F and T, were trained in an instrumental free-choice probabilistic learning task. All procedures were approved by the Institut de Neurosciences de la Timone Ethics Committee (Protocol A2-10-12) and were in accordance with the principles of the European Union Directive 2010/63/EU on the protection of animals used for scientific purposes. The surgically implanted monkeys were head-restrained to allow for stable electrophysiological recordings in the striatum.

Both monkeys were involved in previous experiments studying single-neuron activity in the striatum during performance of a task that involves reaching arm movements to a visual target (Marche et al., 2017; Marche and Apicella, 2021). As shown in **Figure 1A**, the experimental setup consisted of three targets (metal buttons of 10-mm in diameter) aligned horizontally, at the monkey's eye level, on a panel that was placed at a distance of 30 cm in front of the animal. The distance between targets was 10 cm. A two-color (red and green) light-emitting diode (LED) was located below each target. Monkeys were trained to hold a metal bar, located on the lower part of the panel at their waist level, as a starting position for the movement. A tube positioned directly in front of the animal's mouth dispensed small amounts of fruit juice (0.3 ml) as reinforcement. The liquid was delivered through a solenoid valve which made a brief noise whenever it opened, potentially acting as a secondary reinforcer in rewarded trials.

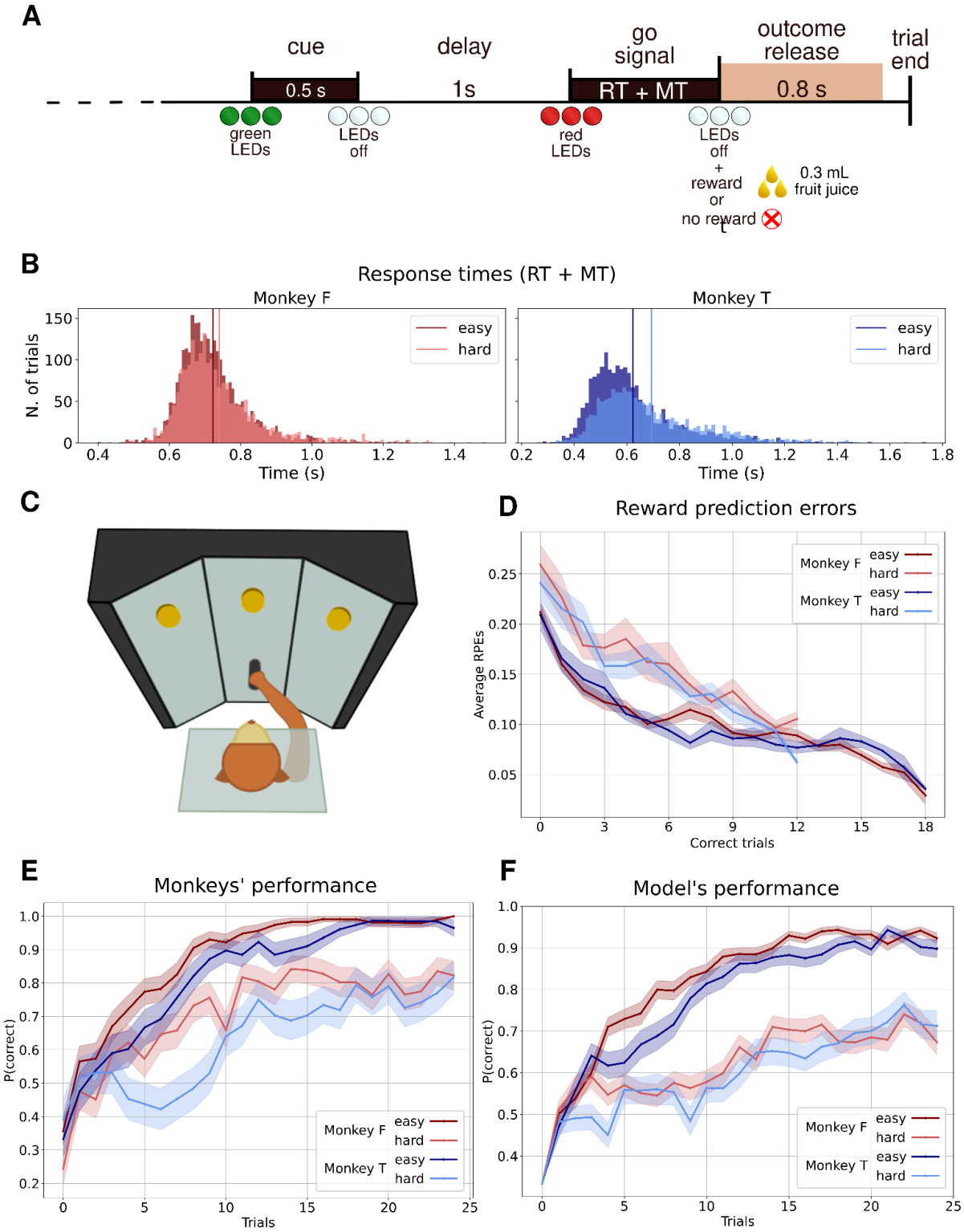

Figure 1. Probabilistic learning task and choice performance. A) Sequence of events inside a single trial. Each trial started with the monkey holding its hand on a metal bar. After a first visual stimulus ('cue' ~~onset~~, green LEDs on) lasting 0.5 s, a second visual stimulus ('go signal', red LEDs on) was presented ~~1 s~~ after 1 s delay ~~the cue offset and instructed~~ to instruct the monkeys to perform a reaching movement to one of the three targets. After a variable delay depending on the reaction time (RT) and the movement time (MT) of the monkeys, on target contact, the go signal was turned off and the monkey immediately received an outcome (reward or not). Correlates of the

RPE signals were examined in an 800 ms period (orange-shaded area) after outcome release. **B)** Histograms representing the distribution of the motor response times (composed of RT + MT) relative to monkey F and monkey T for both Easy and Hard conditions. Vertical lines represent the mean of the distributions. ~~**B**~~**C)** Experimental setup. Monkeys sat in a box with two openings, one for the head and one for their right arm, in front of three target buttons with LEDs, that could be reached with their right hand. An equally reachable metal bar placed under the middle button was used as the starting position of a trial. ~~**C**~~**D)** Evolution of RPE as a function of correct trials. Correct trials are considered as the trials in which the monkeys chose to press the most rewarding button. Data were pooled across blocks for each schedule ("Easy", "Hard") and each monkey. The solid lines and shaded areas correspond to the mean ± standard error of the mean (SEM) of RPEs computed by the Q-learning model. ~~**D**~~**E)** Choice performance computed from monkeys behavior. Data were pooled across blocks for each schedule ("Easy", "Hard") and each monkey. The solid lines and shaded areas correspond to the mean ± standard error of the mean (SEM) of the probability of choosing the most rewarding target as a function of trial number within a block. ~~**E**~~**F)** Choice performance computed from the Q-learning model. Data were pooled across blocks for each schedule ("Easy", "Hard") and each monkey. The solid lines and shaded areas correspond to the mean ± standard error of the mean (SEM) of the probability of choosing the most rewarding target extracted trial by trial from the state-action transition matrix computed by the model.

A trial was initiated when the monkey kept its hand on the metal bar for 1 s, after which all LEDs were lit with a green color for 500 ms ('cue onset' in **Figure 1A**). A fixed delay period of 1 s followed the 'cue offset'. At the end of the delay period, all LEDs turned red ('go signal'), which served as a trigger stimulus for choosing among one of the three targets. Monkeys were trained to reach and touch one of the three possible targets. At target contact, all stimuli turned off and a feedback, constituted solely by the presence or absence of the reward, was provided to the monkeys. Liquid rewards were delivered according to a predefined probabilistic reward schedule, and we kept the reward magnitude constant (0.3 ml) irrespective of the schedule. Regardless of the presence or absence of reward, monkeys had to bring the hand back on the bar to initiate the next trial. A new trial began only if a total of 6 s has elapsed from the initiation of the trial. Trials in which the monkey released the bar before the onset of the go signal were aborted. Trials in which the monkey did not release the bar within a maximum of 1s after trigger onset or in which it did not contact a target within a maximum of 1s after bar release were excluded from subsequent analysis.

Monkeys were trained to perform the task under two probabilistic reward schedules. The two conditions differed in the degree of uncertainty of reward delivery. In the "Easy" condition, the reward probabilities associated with the three targets were (0.7, 0.15, 0.15). In the "Hard" condition, the reward probabilities were (0.5, 0.25, 0.25). During a recording session, the location of the target with the highest reward probability and the probabilistic reward schedule were varied pseudorandomly across blocks of trials. Since no explicit signal informed the monkey which of the targets was the most rewarding, the monkey's behavioral strategy was to learn and ameliorate choices by trial-and-error. Each block lasted a varying number of trials (30 to 80 trials) to prevent anticipation of a block transition by the number of trials. For each trial, we measured the duration of the reaching movement, composed of the reaction time (RT, defined as the time interval between the go signal and the bar release) and the movement time (MT, from the bar release to the target contact), and the chosen target.

2.1.2. Acquisition of neurophysiological data

We used conventional techniques for recording neuronal activity from striatum (Marche et al., 2017; Marche and Apicella, 2021). Monkeys were implanted with a recording chamber

targeting the striatum, centered on the anterior commissure (AC), which allowed vertical access to the putamen and caudate nucleus with custom-made glass-coated tungsten microelectrodes (impedance: 1–2.5 MΩ). The microelectrode was passed inside a stainless steel guide tube lowered through the dura mater and advanced with a manual hydraulic microdrive (MO95, Narishige, Tokyo, Japan). Recordings were made in striatal sites where single-neuron activity was found, and the sites changed from one recording session to another within the limits of the exploration area permitted by the chamber. LFP signals were amplified (x 5000), bandpass filtered (3-150 Hz), and then sampled at 16.6 kHz by using a Power1401 Analog-Digital converter and a multi-channel acquisition software (Spike2, version 7.2; Cambridge Electronic Design, UK).

2.1.3. Histological reconstructions

Recording sites were histologically verified in both animals, using several electrolytic lesion marks in the putamen anterior and posterior to the AC (Marche et al., 2017; Marche and Apicella, 2021). Upon completion of electrophysiological recordings, monkeys were deeply anesthetized by using pentobarbital and perfused with 4% paraformaldehyde. Coronal brain slices (40 μm thickness) containing the striatum were prepared and stained with Cresyl violet to identify the lesion marks. Electrode penetrations were reconstructed in serial sections through the striatum in each monkey.

2.2. Behavioral learning model

In order to model behavioral choices and to estimate the evolution of RPEs during learning, we used a standard modeling approach based on animal associative learning theories (Dickinson, 1980; Wasserman and Miller, 1997). We assumed that probabilistic learning resides in the computation of cue-response-outcome associations, whose strengths depend on the contingency and contiguity of the events (Rescorla, 1991; Dickinson, 1994; Wasserman and Miller, 1997; Balleine and Dickinson, 1998). To quantify the evolution of the associative values and RPEs (i.e., the discrepancy between the observed and predicted outcome), we implemented the Rescorla-Wagner model (1972) as a form of the Q-learning algorithm (Watkins and Dayan, 1992) from reinforcement learning theory (Sutton and Barto, 1998). The Q-learning model has been largely used in previous neuroimaging and neurophysiological studies, and it represents a standard approach for behavioral-modeling for the analyses of neural data (Schultz, 2006; O'Doherty et al., 2007).

Briefly, the Q-learning model updates action values through the Rescorla-Wagner learning rule (1972) expressed by the following equation:

$$Q_a(t + 1) = Q_a(t) + \lambda \cdot \Delta Q \quad (1)$$

where $Q_a(t)$ corresponds to the value of action a = 1, 2, 3 (three possible movements to 3 targets) at trial *t*, and λ is the learning rate (usually ranging from 0 to 1). ΔQ corresponds to the update value, also called Reward Prediction Error (RPE):

$$\Delta Q = RPE = r(t) - Q_a(t) \quad (2)$$

where r models the type of outcome (i.e., r equals 1 for rewards, 0 otherwise). Action values Q_a are then transformed into probabilities according to the softmax equation:

$$P_a(t) = exp(\beta Q_a(t)) \,/\, \sum_a exp(\beta Q_a(t)) \quad (3)$$

The coefficient β is termed the inverse 'temperature': lower β (less than 1) causes all actions to be (nearly) equiprobable, whereas higher β (greater than 1) amplifies the differences in association values. For each block of trials we fitted separately two free variables of the model: the learning rate of the learning rule (λ) and the inverse of the temperature used by the softmax function (β). To do so, we used a grid-search approach to find the best fitting couple of values, varying the value of λ from 0.1 to 1 (in steps of 0.01) and of β from 1 to 10 (in steps of 0.2). We identified the set of parameters that best fitted the behavioral data using the log-likelihood of the probability to make the action performed by the animal, computed as follows:

$$L = \sum_t ln\, P_{chosen}(t) \quad (4)$$

The set of parameters associated with the maximum log-likelihood were used for the estimate of RPEs.

A Q-learning model was fit to the behavioral data of each learning block separately. This produced a set of model parameters (i.e., learning rate λ and inverse 'temperature' β) for each learning block.

2.3. LFP data analysis

2.3.1. Preprocessing of LFP data

LFP signals were preprocessed using a 50Hz notch filter and a band pass filter between 1 and 140 Hz. LFP time series were epoched and aligned on target contact (i.e., outcome onset), termed the outcome period. Visual examination was performed to remove recordings where the LFP activity was contaminated by the spiking activity of surrounding neurons at the sites of LFP recording, despite a low-pass filter being applied on data. Trials with evident electrical artifacts were also discarded. We discarded 28 blocks of trials out of a total of 222 blocks for monkey F and 72 blocks of trials out of 213 for monkey T. In most of the cases, those trials presented a broad-band increase in power visible when computing the time-frequency map. Baseline activity was considered as the LFP data in a time interval from -550 ms to -50 ms relative to cue onset. Filtered LFP signals were epoched into 0.8 s epochs aligned on target contact and downsampled to 1000 Hz for further analysis. Since each block lasted a varying number of trials per block (30 to 80 trials), we considered for subsequent analysis the first 25 trials in each block. This was motivated by the need to have an equal number of trials across blocks. Overall, the final dataset consisted of 194 blocks for monkey F (114 "Easy" + 80 "Hard") and 141 blocks for monkey T (78 "Easy" + 63 "Hard").

2.3.2. Single-trial estimates of LFP power spectra

In order to estimate single-trial and time-frequency representation of LFP power, we used the Morlet wavelet method (Cohen, 1995). Power spectra were computed on 55 frequency steps, logarithmically spaced, in the range between 8Hz to 120Hz, and in a period of time lasting 0.8 s after target contact, corresponding to the outcome period. This temporal window was selected in order to focus on post-outcome relevant signals and to avoid contamination by monkeys' movements (e.g. arm movements) and by sporadic artifacts happening when

the monkey touched the metal bar to return to starting position. The number of cycles used for each band was equal to its frequency divided by 4, in order to obtain wavelets of the same length (i.e., time duration, in this case 250 ms) for each frequency band. We computed the relative change of the time-frequency power of the LFP with respect to the baseline power. With this procedure, we obtained a single-trial time-frequency representation of normalized LFP power for each recording block.

In order to estimate single-trial and band-limited time courses of LFP power, we used the multitaper method based on discrete prolate spheroidal (slepian) sequences (Percival and Walden, 1993; Mitra and Pesaran, 1999). To extract single-trial beta-band power, LFPs time series were multiplied by k orthogonal tapers (k=4) (0.33 s in duration and 12 Hz of frequency resolution), and then Fourier-transformed. The monkeys-specific central frequency (25 and 30 Hz for monkey F and monkey T, respectively) for the beta estimation were established after a statistical analysis performed between time-frequency maps of rewarded and unrewarded trials. Thus, the beta power for monkey F was computed on a frequency range of 19-31Hz, and the beta power for monkey T was computed on a frequency range of 24-36 Hz.

2.3.3. Information theoretical analysis of LFP data

We used information-theoretic metrics to quantify the statistical dependency between the band-limited beta-band power and RPE signals. To this end, we computed the mutual information (MI) between the single-trial and time-resolved LFP power and the behavioral variable. As a reminder, MI is defined as:

$$I(X;Y) = H(X) - H(X|Y) \tag{5}$$

where the variables X and Y represent the trial-by-trial power of the LFP and RPEs, respectively. $H(X)$ is the entropy of X, and $H(X|Y)$is the conditional entropy of X given Y. Entropy estimates were computed using a semi-parametric binning-free Gaussian-Copula approach (Ince et al., 2017). In brief, the GCMI approach exploits the fact that MI is invariant under monotonic transformations of the marginals. This result can be exploited to render the joint distribution of the variables Gaussian by means of local transformations on the marginals, using the so-called Gaussian copula. GCMI therefore requires transforming the X and Y variables so that the marginal distributions are a standard normal. This copula-normalization involves calculating the inverse standard normal cumulative density function (CDF) value of the empirical CDF value of each sample, separately for each input dimension (i.e., sum-rank computation). Then, entropy values can be estimated using a standard covariance-based formula for Gaussian distributed random variables. We also included a parametric bias-correction for the estimate of the entropy values, which is an analytic correction to compensate for the bias due to the estimation of the covariance matrix from limited data (i.e., limited number of trials). In fact, the limited sampling bias is known to affect the estimation of information theoretical measures (Panzeri and Treves, 1996). The Gaussian-Copula Mutual Information (GCMI) is a robust rank-based approach allowing to detect any type of relation as long as this relation is monotone. Since the current analyses involve univariate continuous variables, such information theoretical analysis is equivalent to a Spearman rank correlation approach.

2.4. Statistical analysis

2.4.1. Statistical analysis of behavioral data and model parameters

In order to quantify the evolution of learning during each learning block, we computed the probability of choosing the most rewarding target as a function of trial number. To do so, we pooled data across blocks for both schedules (“Easy”, “Hard”) and averaged the binary outcomes across blocks. In order to quantify potential differences in learning processes across conditions and animals, we performed a two-way ANOVA on each of the learning model parameters (i.e., learning rate λ and inverse ‘temperature’ β). The first factor was the monkey (T and F) and the second was the experimental condition (“Easy” and “Hard”). The analysis of learning rate λ was meant to assess differences in learning speed across monkeys and conditions, whereas the analysis of the inverse ‘temperature’ β assessed differences in behavioral strategy.

We then investigated the relation between RPEs and the learning dynamics within each block. In particular, we focused on positive RPEs observed after the selection of the most rewarding target (i.e., the “correct” action). The rationale was to investigate the relation between learning dynamics and RPE signals that drive the update in action values, thus positive RPEs. We expected to observe higher values of RPEs early during learning and smaller RPEs later during learning. In addition, we expected to observe a statistical significant difference among conditions. We therefore analyzed exclusively trials in which the monkey was rewarded after the selection of the correct (most rewarding) target. For each trial, we extracted the RPE signal and the trial index (i.e., ranging from 1 to 25 within a learning block). We then sorted trials according to the RPE value and created four equally-sized groups according to the percentile RPEs: i) below the 25th percentile; ii) from the 25th to 50th percentile; iii) from the 50th to 75th percentile, iv) above the 75th percentile. For each group of trials, we calculated the average trial index defined as the mean trial index. Such analysis was separately performed for each monkey and experimental condition. Statistical analysis was performed by means of a two-way ANOVA, where the first factor was the percentile range (4 levels) and the second factor was the experimental condition (“Easy” and “Hard”).

2.4.2. Statistical analysis of LFP data

Two types of statistical analyses were performed on LFP data. The first aimed at finding the frequency range and peak at which a significant outcome-related modulation (i.e., difference between rewarded and non-rewarded trials) was observed in the LFP signals. To do so, for each monkey, we performed a two-tailed t-test on the single-trial time-frequency representations, and we contrasted rewarded and unrewarded trials. The resulting p-values were Bonferroni corrected across the total number of points composing each time-frequency map. For each monkey, we found a peak of significance related to the beta-band activity, which was used for the band-limited analyses of LFP data.

For the statistical analysis of RPE-related modulations in LFP power, as assessed by means of Gaussian-Copula Mutual Information (GCMI), we used a group-level approach based on non-parametric permutations (Combrisson et al., 2022). The time-resolved GCMI was estimated between the LFP power and the behavioral variable (RPE) by concatenating trials across blocks for each electrode. For statistical analyses, we adopted a fixed-effect model across blocks of trials for each monkey (respectively 194 and 141 blocks for monkey

F and T). By estimating the effect size across blocks, we improved the statistical power and the overall signal-to-noise ratio at the cost of ignoring the block-to-block random variations. To do so, we generated 1000 permutations by randomly shuffling the vector of RPE, allowing us to sample the distribution of MI reachable by chance (Combrisson and Jerbi, 2015). To correct for multiple comparisons, we used a cluster-based approach with clusters detected across time points (Maris and Oostenveld, 2007). The cluster-forming threshold was defined as the 95th percentile of all of the permutations (i.e., across time points and electrodes). This threshold was then used to form the clusters on the true MI and on the permutations. Finally, the corrected p-values were inferred as the proportion of the maximum of the cluster-mass detected from the permutations exceeding the true estimation of MI.

As a control analysis, we fitted a multiple linear regression model estimating the relationship between the beta-band LFP power as dependent variable and six independent variables, three that are classically considered associated with outcome-related processes, i.e. the reward, RPE and the absolute value of the RPE (absRPE), and three associated with action-related variables, i.e. reaction times (RT), movement times (MT) and the chosen action (Action). A multiple linear regression model was fitted to each recording block and group-level analysis was performed on the single-block beta coefficients using a two tailed t-test.

2.5. Analysis of anatomical specificity of RPE signals in striatal territories

We next investigated whether the encoding of RPEs by beta-band LFP power differentially recruited the sensorimotor, associative and limbic territories of the striatum. To do so, we performed RPE-related analyses on LFP power modulations in subgroups of recordings associated with different striatal territories. The localization of the recording site within the striatum was done according to previous studies (Parent, 1990) and based on the stereotaxic atlas of Paxinos et al. (2008). The anterior commissure was used as a landmark to separate the associative and limbic striatum (dorsal and ventral parts of the precommissural caudate nucleus and putamen, respectively) from the motor striatum (dorsal part of the postcommissural putamen). For each monkey, the center of the recording chamber corresponded to the location of the anterior commissure. Each electrode track was performed using specified XY coordinates (AP, ML), referenced to the central position of the chamber, and the depth of each recording site was referenced to the tip of the guide cannula inserted into the brain, above the striatum. We measured the antero-posterior (AP, X-axis) and medio-lateral positions (ML, Y-axis) from the center of the recording chamber, and the dorsoventral position from the tip of the cannula (depth). Each recording session was therefore labeled as located in either the sensorimotor, associative and limbic striatum. For monkey F in “Easy” condition, we analyzed 30 blocks (855 trials) in the limbic striatum, 42 blocks (1200 trials) in the associative striatum, and 42 blocks (1181 trials) in the motor striatum, while in the “Hard” condition we analyzed 20 blocks (583 trials) in the limbic striatum, 30 blocks (921 trials) in the associative striatum, and 30 blocks (986 trials) in the motor striatum. For monkey T in “Easy” condition we analyzed 27 blocks (653 trials) in the limbic striatum, 24 blocks (533 trials) in the associative striatum, and 27 blocks (681 trials) in the motor striatum, while in the “Hard” condition we analyzed 23 blocks (588 trials) in the limbic striatum, 23 blocks (708 trials) in the associative striatum, and 17 blocks (523 trials) in the motor striatum.

In order to investigate the presence of functional gradients across regions of the striatum and local selectivities of RPE-related modulations in beta-band LPF power, we subdivided recording sessions into different groups according to their spatial location. To do so, we employed a K-means algorithm applied to the 3-dimensional spatial coordinates (AP, ML and depth) of the recording sites within each territory (sensorimotor, associative and limbic). The K-means algorithm allows a uniform repartition of the recording sites according to their 3D spatial coordinates and proximity. The number of clusters in each territory was set to achieve an optimal trade-off between a fine spatial selectivity (i.e., maximizing the number of clusters) and the amount of data (i.e., number of learning blocks and trials within each cluster). Thus, we set the number of clusters equal to six for each striatal territory (sensorimotor, associative and limbic), obtaining a total of eighteen spatial clusters across the sampled striatal regions. Finally, we computed the distance between the centroid of each cluster and a reference point set as the highest and most rear coordinates across all recording sites for each of the two monkeys. We then re-reference the subcluster positions with respect to a rostro-ventral to caudo-dorsal axis. We used such distance values and the average MI computed across the blocks of trials belonging to each cluster to study the distribution of RPE related information across different striatal territories.

2.6. Software

All data analyses were performed with subroutines written in Python (version 3.6). The preprocessing and spectral analysis of LFP data was performed with neo (version 0.8.0) (Garcia et al., 2014) and MNE (version 0.21) ((Gramfort, 2013). Data management and storage was performed using pandas (version 1.1.5) (McKinney, 2010) and xarray (version 0.16.2) (Hoyer and Hamman, 2017). Analysis and statistics on behavioral data were performed using scikit-learn (version 0.23.1) (Pedregosa et al., 2011) and statsmodels (version 0.12.2) (Seabold and Perktold, 2010). The statistical analysis of LFP data was performed using Frites (version 0.3.8) (Combrisson et al., 2022). Figure production was performed using matplotlib (version 3.3.4) (Hunter, 2007) and plotly (version 4.14.3) (Plotly Technologies Inc., 2015).

3. Results

3.1. Behavioral results

The evolution of behavioral performances shows that both monkeys learned by trial-and-error which target was most rewarding over the course of each block of trials. Each block was characterized by an initial exploratory phase that allowed monkeys to find the most rewarding action, followed by a phase in which monkeys preferentially chose the most rewarding target until the end of the block. In order to quantify behavioral performance across monkeys, we aligned all blocks and computed the probability of choosing the most rewarding target among the three options. As we can see from the progression of the curves in **Figure 1D**, approximately 15 to 20 trials were sufficient for both monkeys to identify the position of the most rewarding target for both the conditions. Monkeys had a tendency to learn quicker and chose more often the most rewarding target in the “Easy” condition than in the “Hard” one (**Figure 1D**). Indeed, we computed the average λ (learning rate) values obtained by model fitting for both monkeys and conditions: average λ values for monkey F were 0.292 and 0.342 respectively for the “Easy” and “Hard” conditions, respectively. The average λ values for monkey T are 0.288 and 0.334 for the “Easy” and “Hard” conditions, respectively. The average of the β values (inverse of the softmax temperature) were 9.765 and 9.398 for monkey F ~~(“Easy”/“Hard”)~~ and to 9.554 and 9.168 for monkey T ~~(“Easy”/“Hard”)~~ for the “Easy” and “Hard” conditions, respectively. As mentioned in the Methods section, the range of β values used in the grid-search algorithm to find the best set of parameters was set in between 1 and 10. In a control analysis, we tested the reliability of the fitting algorithm and parameter space. We thus fitted the model using a more sophisticated algorithm (i.e., the truncated Newton algorithm or TNC) for likelihood minimization and we increased the range of possible β values (from 1 to 10000). Although the model's performance in fitting monkeys’ behavior was ameliorated, we observed that the Pearson correlation between single-trial RPEs computed with the former and the latter fitted parameters were highly correlated. On average, only 3% of sessions displayed a Pearson correlation less than 0.95. We additionally repeated the mutual information analysis shown in Figure 3B with the new RPE values, and we obtained nearly identical MI values showing the same time course (data now shown).

In order to quantify differences in learning rate and behavioral strategies across conditions and monkeys, we performed a two-way ANOVA on the across-blocks model parameters (λ and β). The first factor was the monkey (T and F) and the second was the experimental condition (“Easy” and “Hard”). Significant differences both in λ and β values were observed across conditions (λ p-value=0.004, β p-value=0.005). No significant effect was observed across monkeys or at the level of the interaction between the two factors (p-values > 0.05).

We then investigated the relation between RPEs and the learning dynamics within each block. To do so, we performed a two-way ANOVA on the trial indices within each block, where the first factor was the RPE percentile level (4 ranges) and the second factor was the experimental condition (“Easy” and “Hard”). **Figure 1C** shows that the higher values of RPEs are associated with lower average trial number, whereas lower values of RPEs are associated with higher number of rewards on correct trials.

The overall number of rewarded correct trials related to RPEs percentiles is lower in the “Hard” condition with respect to the “Easy” condition because of the differences in reward

schedules. A two-way ANOVA analysis confirmed the significance of this relation for both the monkeys (**Table 1**).

	Monkey F			Monkey T		
	F-statistic	**Degrees of freedom**	**p-values**	**F-statistic**	**Degrees of freedom**	**p-values**
RPE percentiles	205.399	3.0	<0.001	132.624	3.0	<0.001
Condition (easy / hard)	230.170	1.0	<0.001	152.167	1.0	<0.001
RPE percentiles : Condition	12.624	3.0	<0.001	7.117	3.0	<0.001

Table 1. Summary statistics of the distribution of rewarded correct trials in relation with RPEs percentiles ranges and the two task conditions (“Easy” / “Hard”) (see **Figure 1C**). p-values are obtained with a two-way ANOVA test.

3.2. Reward modulates beta-band LFP power

We then investigated whether modulations in striatal beta-band LFP activity differed among rewarded and unrewarded trials. To do so, we computed for each learning block the average time-frequency power for all rewarded and unrewarded trials and the difference between the two, until 0.8 s after the target contact and outcome presentation and in a range of frequencies from 8 to 51 Hz (**Figure 2B**). We performed a two-sided t-test analysis across the two types of outcomes, and then we Bonferroni-corrected the p-values with respect to the total number of points in the time-frequency representation. Highly significant portions of the time-frequency representation displaying outcome-related modulations were observed for both monkeys in the beta band and around 0.4s after outcome presentation (**Figure 2C**). This analysis allowed us to identify the peak frequency in each beta band displaying the strongest modulation for subsequent band-limited analyses. The central frequency was 25 Hz for monkey F and 30 Hz for monkey T.

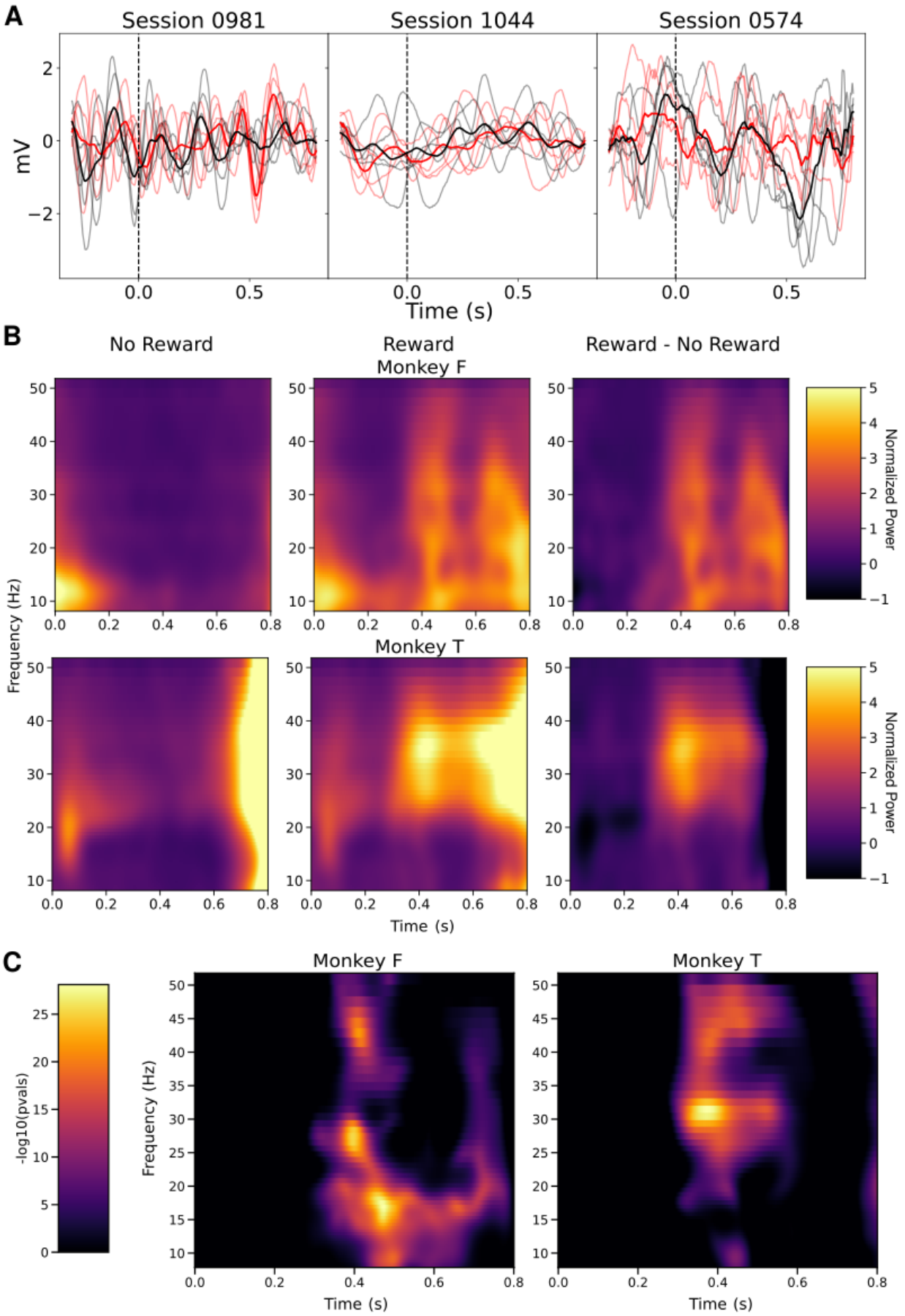
A
Session 0981
Session 1044
Session 0574
mV
Time (s)
B
No Reward
Reward
Reward - No Reward
Monkey F
Monkey T
Frequency (Hz)
Normalized Power
Time (s)
C
Monkey F
Monkey T
-log10(pvals)
Frequency (Hz)
Time (s)

Figure 2. Outcome-related modulations in LFP power. **A)** Exemplar sessions depicting epoched LFPs (five rewarded and five unrewarded trials each). ~~from Monkey F in both "Easy" (left) and "Hard" (right) conditions~~. LFPs were filtered between 1 and 140Hz. Light red lines represent rewarded trials while light black lines represent unrewarded trials. The dashed vertical black line coincides with the target contact and outcome release ~~delivery~~, and is the temporal point on which data are aligned. The two plain red and black lines ~~small inserts in the bottom-left part of each plot~~ represents the corresponding averages. **B)** Time-frequency maps of Monkey F (top) and Monkey T (bottom), averaged across trials within both the task conditions, grouped by the outcome (No reward, Reward) and the subtraction between the two (Reward - No Reward). The time window from 0 s to 0.8 seconds corresponds to the outcome period (orange-shaded area in **Figure 1A**), selected to avoid the contamination from relevant movements (e.g. arm movements) **C)** The statistical analysis of LFP power modulations contrasting rewarded and unrewarded trials displayed significant effects in the beta band. The color code is in the -log10(p-values) scale.

3.3. Beta-band LFP correlates of RPEs

One of the main goals of the study was to investigate the relation between beta band power modulations and RPEs. **Figure 3A** shows that in the limbic striatum, a striatal territory in which we expected to find a strong correlation between neural activity and RPE signals, the relation between the average beta power integrated over a time window of 0.2 - 0.8 seconds and the evolution of RPE values along trials highlights a nonlinear pattern for each of the two monkeys. In order to statistically quantify the relation between outcome-specific modulations in beta-band power and RPE signals over the entire dataset, we computed the mutual information (MI) between evolution of RPEs and beta-band power of the LFP activity across trials in a time-resolved manner. Statistical analysis was performed across all sessions using cluster-based statistics combined with permutation tests. In preliminary analyses, we computed the MI between the beta-band activity and time-resolved LFP power separately for the "Easy" and "Hard" conditions. Since no significant difference in MI was found across conditions (result not shown), we concatenated trials for the two conditions in subsequent analyses. **Figure 3B** shows the time-course of the MI between RPEs and beta-band LFP power. In both monkeys, the time-course of MI increased around 200 ms, peaked around 450 ms after outcome onset and lasted a total of approximately 550 ms. Significant temporal clusters ($p<0.05$) obtained by means of cluster-based statistics and permutation tests are represented in the plot by the continuous line (see details about the statistical analyses in the **Materials and Methods** section). Overall, these results show that beta-band power modulations in the striatum are differentially modulated by the presence or absence of reward (**Figure 2**) and encodes RPE signals (**Figure 3**).

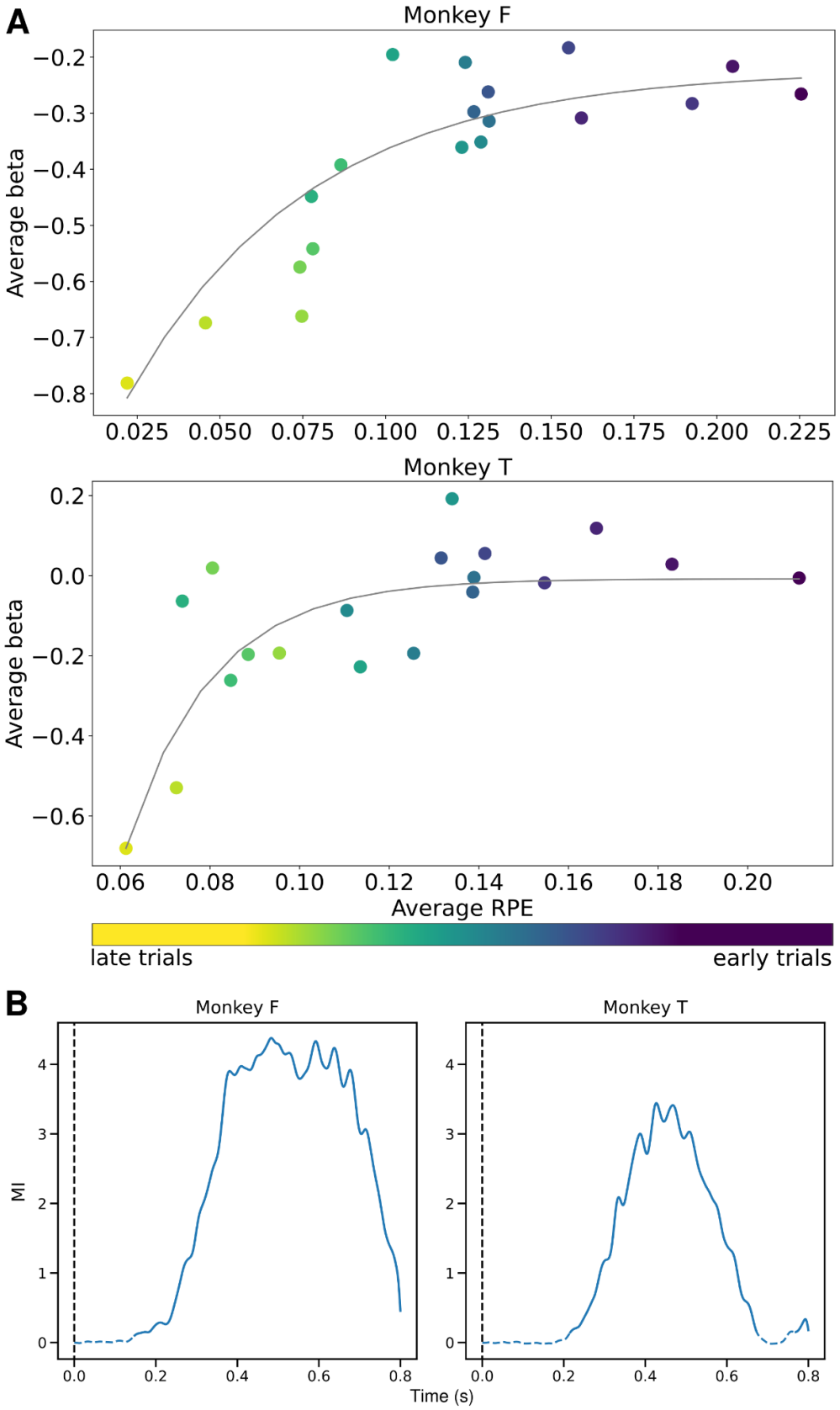

Figure 3. Relation between RPEs and beta power. A) Relation between the averaged normalized beta power over a time window of 0.2-0.8 seconds and average RPEs across trials in the limbic striatum of Monkey F (top) and Monkey T (bottom). Dots' color fading from yellow to blue represents the passage from early trials to late trials. The negative values of the averaged beta

power are the results of the normalization over the baseline. **B)** Mutual Information (MI) between beta-band LFP power and RPE. The dashed vertical line represents the target contact time on which data are aligned. The dashed blue lines represent non significant values ($p \geq 0.05$) of MI, while the continuous ones represent significant values ($p < 0.05$). The chosen time window reflects the outcome period, with time 0 corresponding to the target contact and outcome delivery.

In order to assess the potential contribution of additional task variables (i.e., outcome types and RPEs) to trial-by-trial LFP power modulations, we fitted a multiple linear regression model estimating the relationship between the beta-band LFP power as dependent variable and the six independent variables already mentioned in Methods, section 2.4.2: i) reward; ii) RPE; iii) absRPE; iv)RT; v) MT and vi) Action. **Table 2** shows the results of the statistical analyses. The only regressor which displayed a significant contribution in both monkeys to the beta-band LFP power was the RPE. Additionally, we found a significant contribution of the beta band average activity in the encoding of reward and absRPE in Monkey F, and of RT and MT in Monkey T. Due to the lack of reproducibility across monkeys, subsequent analyses were focused on RPEs correlates only.

	Monkey F		Monkey T	
	F-value	**p-value**	**F-value**	**p-value**
Reward	7.48	0.006	0	0.99
RPE	5.668	0.017	5.591	0.018
absRPE	5.735	0.017	0	0.998
RT	3.691	0.055	5.641	0,018
MT	0.61	0.435	12.914	0
Action	1.489	0.226	1.239	0.29

Table 2: ANOVA summary table resuming the results obtained from the multiple regression analysis between average beta-band LFP oscillations in the outcome period and six behavioral regressors.

3.4. Anatomo-functional correlates of RPEs in monkey striatum

We next investigated whether the encoding of RPEs by beta-band LFP power differentially recruited the sensorimotor, associative and limbic territories of the striatum. Indeed, the neurophysiological recordings were made in all territories of the striatum, including sensorimotor, associative and limbic portions. **Figure 4** illustrates the spatial distribution of striatal recording sites in monkey F, as verified by histological analysis. The

neuronal sample was taken from approximately the same striatal regions in monkey T (not shown).

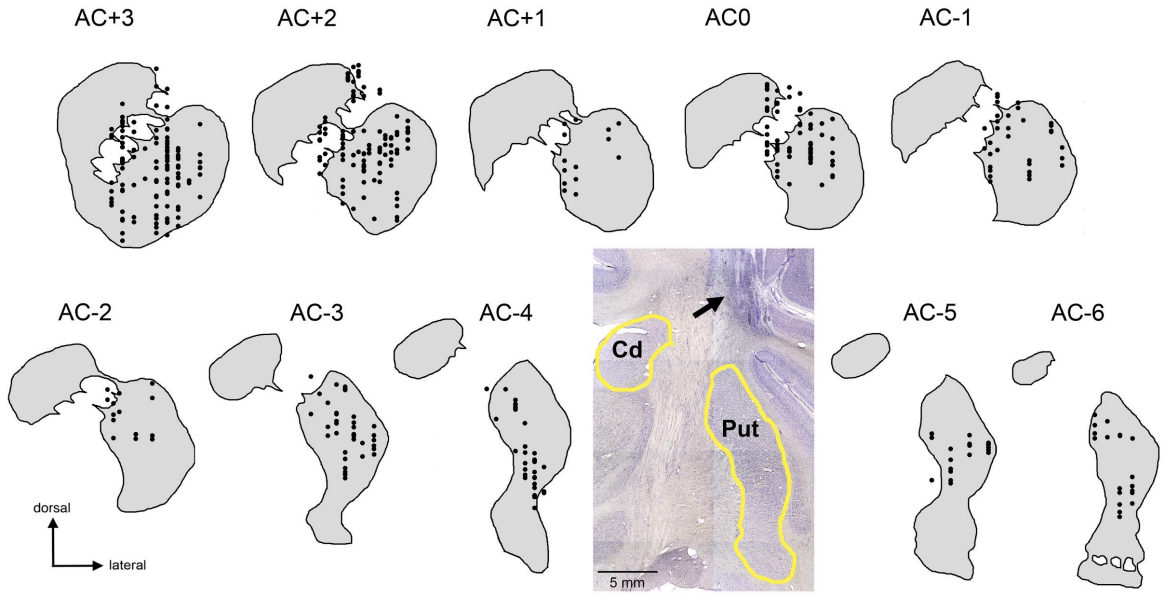

Figure 4. Positions of all striatal recording sites in monkey F. Each dot corresponds to a single LFP recording site. Coronal sections are labeled in rostrocaudal stereotaxic planes according to distances from the anterior commissure (AC) used as a reference landmark. The inset shows a photomicrograph of a coronal section stained with Cresyl violet at the level of the posterior putamen (i.e., sensorimotor striatum) with visible traces of electrode tracks (arrow) above the putamen. Cd, caudate nucleus; Put, putamen.

Each recording session involved a single electrode recording and sampled the striatum at a single position. To investigate the spatial distribution of the RPE-related modulations, we first labeled the recording sessions into different striatal territories (sensorimotor, associative and limbic). Then, we applied the K-means algorithm to the 3-dimensional spatial coordinates (AP, ML and depth) of the recording sites to obtain a total of eighteen spatial clusters. **Figure 5A** shows the cluster's position relative to the AP position (x axis) and the depth (y axis). The clusters' centers are numbered following the ascending values of the average MI computed for each cluster, split up following the territory division (represented by the colors). In order to study the contribution of each sub-cluster in the encoding of RPEs, we computed the RPE-related MI time courses by grouping all recordings within a given cluster. **Figure 5B** shows the results of our analyses. Each of the three rows correspond to one of the three striatal territories, limbic (red curves), associative (blu) and motor (green) striatum, respectively, while each column corresponds to an anatomical subregion. We observed that the amount of RPE-related MI was higher in the limbic striatum, then gradually decreased towards the associative and motor territories, as shown in **Figure 5B**, in which we can observe the number of significant clusters detected across the striatum. As in **Figure 3**, dashed lines correspond to non-significant time intervals, while full lines correspond to significant temporal clusters.

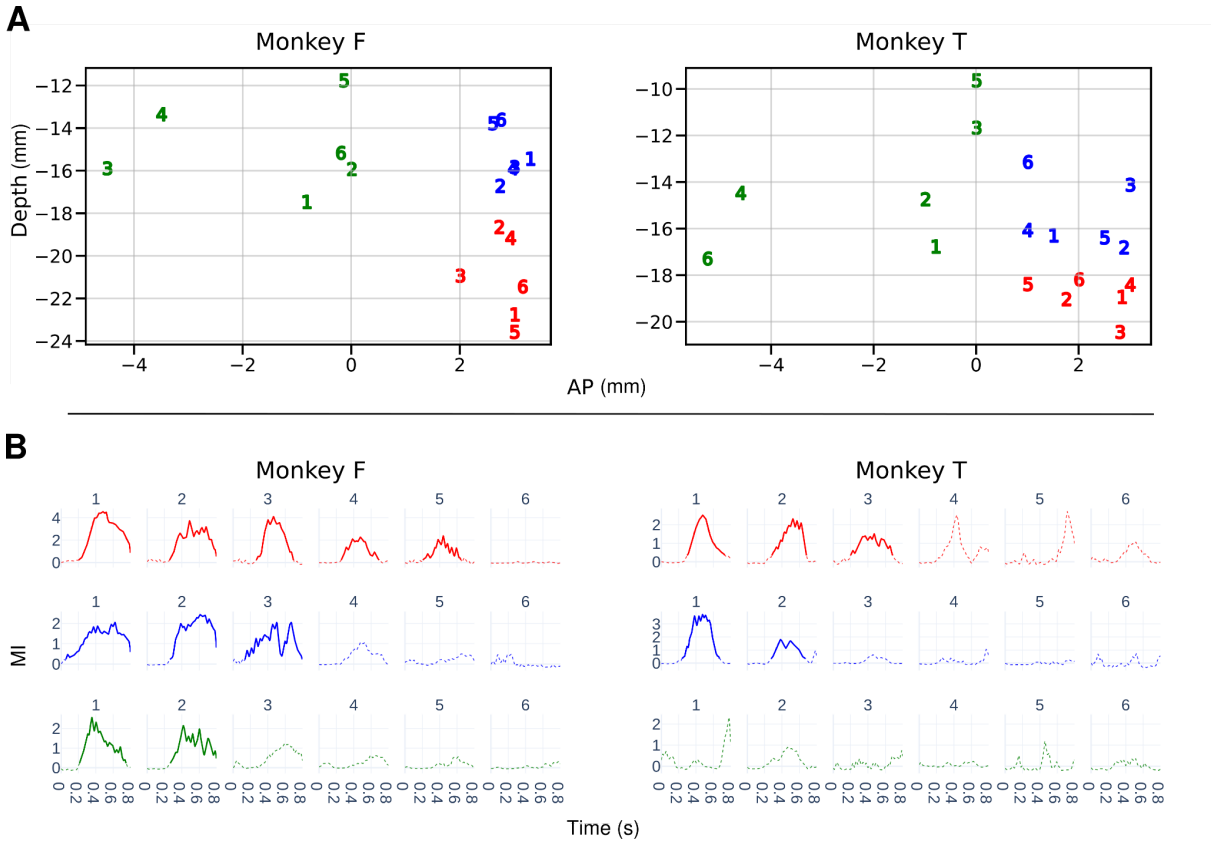

Figure 5. Anatomo-functional distribution of RPE-related beta-band LFP modulations. A) two-dimensional spatial positions of recording sites clusters' centroids, for each monkey. Clusters are represented along their antero-posterior (AP) position and depth of the recording site. Digit numbers are labels that are used in panels B. Digit color corresponds to a striatal region (green for the motor striatum, blue for the associative and red for the limbic parts). **B)** MI computed in each of the clusters, separately for each monkey. Digit labels and digit colors are the same as in A. Dashed and continuous lines represent non-significant and significant values of MI, respectively.

We then assessed whether the effect size in MI about the RPE displayed a spatial organization across the striatum. To do so, we defined a rostro-caudal to dorso-ventral axis by taking the highest and the most posterior among electrodes' positions to define a referential point in space for each of the two monkeys. Then, we computed the Euclidean distance between this reference point and each cluster center, which allowed us to investigate the possible presence of a statistical relation between clusters' positions and functional effects (MI values). As shown in **Figure 6**, we found an increase in RPE information together with the distance from the referential point, toward the rostral-ventral striatum, suggesting a linear progression over distance. To quantify such progression, we performed a linear regression analysis between the distance and the average MI of each cluster. We observed a positive correlation suggesting that the rostro-ventral part of the striatum carries more information about the RPE, and that this information fades toward the caudo-dorsal part of the striatum. Linear correlation analysis revealed a significant and positive correlation (p-values < 0.05) for both monkeys (**Figure 6**). The linear regression with the F-statistic associated p-values are associated with R^2 values of 0.509 (Monkey F) and 0.611 (Monkey T). To summarize in other words, this result indicates that the amount of

information about RPE signals follows an anatomical gradient, showing higher values in the rostro-ventral part of the striatum and gradual decrease towards the most dorso-caudal part.

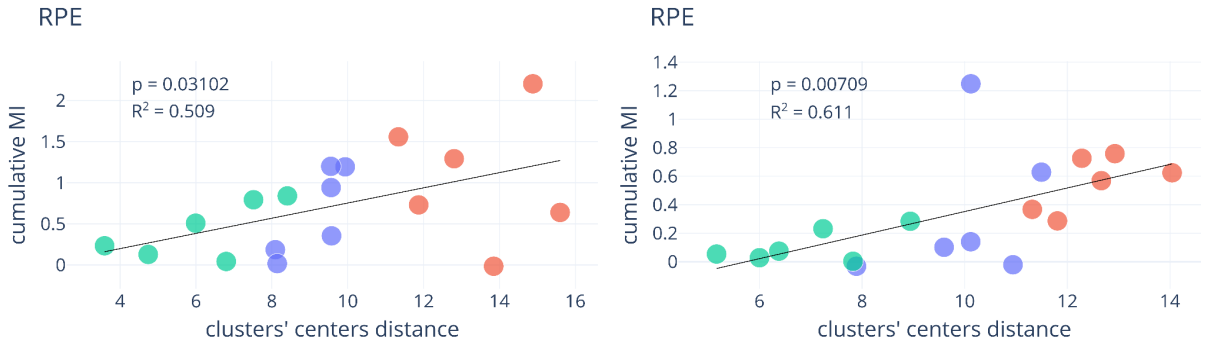

Figure 6. Striatal gradient of the total RPE beta-band MI. Each point represents a cluster of recording sites and colors associated with the three striatal territories (green for motor, blue for associative and red for limbic territories). The Y-axis reflects the cumulative MI calculated over the outcome time interval. The X-axis reflects the clusters' center distance with respect to the rostro-ventral to caudo-dorsal reference. Such reference was computed by taking the AP coordinate of the most posterior recording site and the Depth coordinate of the higher recording site of each monkey.

4. Discussion **(1481/1500 words maX)**

Two main aspects of the functional organization of the striatum emerge from the present study: (1) changes in LFP beta-band oscillations encoding RPE signals (i.e., the difference between expected and actual outcomes) are observed in the striatum; (2) the encoding of RPE is dependent on the striatal region following a rostro-caudal to dorso-ventral gradient, with a maximum in the ventral part of the anterior striatum. These data highlight a relationship of beta oscillatory activity in the striatum to non-motor aspects of behavior, such as the signaling of reward information, and distinct contributions for striatal regions in the evaluation of reward based action outcomes.

4.1 Role of striatal beta oscillations in outcome evaluation

A key finding in our study is the occurrence of LFP beta oscillations during the outcome period of the task that may play a role in evaluative processing after action choice (i.e., presence or absence of reward). Our analysis suggests that RPE signals are a relevant variable influencing striatal LFP beta oscillations, this trend being present in data from every striatal region.

To our knowledge, this is the first report to suggest that striatal beta oscillations play a role in RPE encoding. Indeed, beta-band oscillations in the basal ganglia have been mostly associated with motor control. Numerous studies in humans and animals have provided evidence that an increased beta oscillatory activity within basal ganglia circuitry occurs with

an impaired dopaminergic transmission and the expression of motor deficits observed in humans with Parkinson's disease (Brown, 2007; Jenkinson and Brown, 2011).

Beta oscillations have also been reported in the striatal LFP activity of normal animals, both rodents and monkeys, during specific phases of behavioral tasks (Berke et al., 2004; Courtemanche et al., 2003; Leventhal et al., 2012; Schmidt et al., 2013; Bartolo et al., 2014), but the functional significance of such oscillatory activities is still under debate. In particular, despite the proposed role of the striatum in action valuation and reward-driven learning, few studies have specifically investigated whether striatal beta oscillations can be associated with reward processing (Howe et al., 2011; Leventhal et al., 2012; Münte et al., 2017; Schwerdt et al., 2020). For example, the work of Leventhal et al. (2012) has shown that beta-band oscillations are associated with cue utilization in rats' striatum. The study used four different variants of the classic Go-NoGo task and reported a whole-striatum and non lateralized event-related synchronization (ERS) in the beta-band associated with the cue. Furthermore, these modulations were not linked to motor initiation or suppression. The relevant feature that should follow the cue to produce a beta ERS is the presence of the reward. Overall, these studies suggest that cue-related beta-band power modulations play a role in 'anticipating' the reward occurrence. Similarly, our result shown in **Figure 2** suggests that striatal beta-band plays an important role in outcome processing and not only in anticipation.

4.2 Reward prediction error encoding in the striatum

The role of midbrain dopamine neurons in RPE encoding is well established (Fiorillo et al., 2003; Abler et al., 2006; Bray and O'Doherty, 2007; Fujiyama et al., 2015). Animal electrophysiology and human neuroimaging have provided extensive evidence of RPE-related activity in the striatum (Apicella et al., 2009; Roesch et al., 2009; Oyama et al., 2010; Asaad and Eskandar, 2011; Stalnaker et al., 2012), which is the main target structure of ascending dopamine projections from neurons located in the substantia nigra *pars compacta*. RPE is essential for adaptive behavior in order to avoid non-rewarding actions and exploit the rewarding ones, by improving the predictions about future outcomes (O'Doherty et al., 2017), playing a crucial role in the acquisition of new learned behaviors (Ressler, 2004; O'Doherty, 2007; Keramati et al., 2011; Nonomura et al., 2018). From our work (**Figure 3**), a significant increment of mutual information between the beta-band and the RPE is detected in both monkeys. To interpret this result, we should consider that the MI between two variables can be considered as an index of covariation. Thus, the effect size and an increment in MI corresponds to a strong covariation between the across-trial evolution of the beta-oscillations power and the RPE. Therefore, the striatum can have a major role in encoding and transmission of RPE signals across different functional regions.

More studies about the transmission of RPE signals both intra-striatum and across the striato-cortical network are needed in order to better understand the time course, the localization and the behavioral salience of this signal, so important for the regulation of higher cognitive processes. Finally, we cannot exclude that additional aspects of information processing during the outcome period of the choice task, such as return movements to the resting bar or the experience during reward consumption (sensory pleasure or mouth movements) contribute to the modulations in striatal beta activity. Additional studies are necessary to disambiguate the affective, motor, or cognitive origin of changes in beta oscillations at the end of the trial in our task.

4.3 Functional parcellation of the striatum

Different regions of the primate striatum are assumed to serve different functions, with the dorsal part, including both the caudate nucleus and putamen, involved in cognition and sensorimotor processing, and the rostro-ventral part most closely implicated in reward and motivation (Apicella et al., 1991; Fiorillo et al., 2003; Marchand et al., 2008; Brovelli et al., 2011; Pennartz et al., 2011; Schultz, 2016a, 2016b; Han et al., 2021). We therefore tested such a hypothesis and we investigated LFP activity over the whole striatum searching for differential functional selectivities for action's outcome encoding (**Figure 4**). Indeed, we found that spatially-distant clusters of recording sites differentially responded to action's outcomes (i.e., for rewarded and non-rewarded trials) and differentially encoded RPEs (**Figure 5**). To better understand the spatial organization of the beta-band correlates of RPEs at these sites, we analyzed the relation between the total MI between beta-band LFP power and RPEs, and their relative position along the rostro-caudal and ventro-dorsal axes of the striatum (**Figure 6**). We chose to form clusters that comply with the classic subdivision of the primate striatum into three functional domains, based on the segregation of inputs from cortical and limbic regions (Parent and Hazrati, 1995; Haber, 2003; Jahanshahi et al., 2015).

Several lines of evidence point to a major involvement of the anterior-ventral part of the striatum, including the nucleus accumbens, in processing reward-related information (Apicella et al., 1991; O'Doherty, 2004; Schultz, 2016c). Our results indicate that the information about RPE is distributed in all striatal regions. Nevertheless, we observed a gradient across the striatum, with stronger RPE signals located in the ventral part of the anterior striatum. This novel result is in line with neuroimaging studies in humans highlighting the role of the ventral striatum in the computation of RPE (Abler et al., 2006; Bray and O'Doherty, 2007; Schultz, 2016a; Calderon et al., 2021). Striatal fMRI activity has also been involved in a broad range of functions carried out by parallel organized fronto-striatal pathways (Alexander et al., 1986), spanning from RPE signaling to cognitive control (Mestres-Missé et al., 2012; Vogelsang and D'Esposito, 2018; Alberquilla et al., 2020; Han et al., 2021). It is assumed that RPE signals are needed to update the inner model of action values in response to a particular state, and those values are retained in short term memory in order to plan future actions in a goal-directed way. The distributed RPE information observed in the current study is therefore consistent with the idea that RPEs are important signals that are forwarded to the limbic, associative, and motor networks to influence neural mechanisms that mediate the ability to make value-guided decisions (Silvetti et al., 2014; Schultz, 2016b). Moreover, our results are in line with anatomical studies in monkeys that revealed a topographic organization of connections between midbrain DA neurons and striatal regions that subserves a mechanism by which ascending dopaminergic projections can direct information flow from ventral to more dorsal regions in the striatum (Haber et al., 2000).

4.4 Potential origin of beta-band RPE signals in the striatum

It is generally assumed that LFP oscillations are driven by fluctuations in the excitability of populations of neurons within the recorded region, under the influence of local processing and incoming afferents from other regions (Buzsáki et al., 2012). In our present study, we exclusively analyzed the local relation between the beta power and the RPE in the striatum using a single electrode design. Single-electrode recordings do not allow to precisely assert if, and to what extent, the recorded local activity can be affected by volume conduction

phenomena from afferent distant sources (e.g., cortex and/or thalamus). Indeed, further work is needed to disentangle whether the observed RPE-related modulations are due to local changes in neuronal synchronization, changes in the size of the engaged population, or whether they emerge from coordination phenomena that involve a large-scale brain network and across-brain synchronization processes.

Moreover, it is well established that beta oscillations supports the large-scale coordination across multiple cortical regions involved in different functions, such as sensorimotor integration (Brovelli et al., 2004; Kilavik et al., 2013), visual perception (Vezoli et al., 2021), and working memory (Salazar et al., 2012; Rezayat et al., 2021). Neurophysiological studies in behaving animals have shown that the spiking activity of striatal output neurons and specific interneuron types can be related to beta oscillations in the LFP, raising the possibility that local processing likely contribute to the oscillatory activity in the beta range (Courtemanche et al., 2003; Howe et al., 2011). In addition, it has been demonstrated that cholinergic interneurons in the rodent striatum play a causal role in the generation of beta oscillations in cortico-striatal circuits (Kondabolu et al., 2016).

We suggest that the observed relation between the RPE and striatal beta oscillations is the result of internal striatal computations driven by the dopaminergic system, and involving a larger network supporting learning processes, including additional subcortical and cortical areas (e.g., prefrontal cortex). Overall, our results are in line with the idea that the RPE signals carry crucial information for behavioral update that propagates across different brain regions of the limbic, associative and sensorimotor fronto-striatal circuits (Silvetti et al., 2014; Schultz, 2016b).

Conclusion

To conclude, our study provides new evidence that changes in beta-band LFP oscillations may reflect the encoding of RPEs defined in reinforcement learning models. We observed that RPE-related modulations in LFP power were dominant in the rostro-ventral rather than the caudo-dorsal striatum, supporting the notion of a prominent role for the limbic part of the striatum in evaluative processing useful for future actions. Based on our mapping of the spatial organization of oscillatory beta activity in the striatum, we propose that the RPE encoding can occur first in the ventral region and then spreads over the dorsal region. This finding may be of clinical importance as it is known that dorsal and ventral parts of the striatum are differentially involved in neuropsychiatric diseases, with dorsal striatal circuits mainly related to motor and cognitive disorders, whereas ventral striatal circuits are involved rather in the expression of affective disorders and compulsive behaviors.

Bibliography

Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006) Prediction error as a linear function of reward probability is coded in human nucleus accumbens. NeuroImage 31:790–795.

Alberquilla S, Gonzalez-Granillo A, Martín ED, Moratalla R (2020) Dopamine regulates spine density in striatal projection neurons in a concentration-dependent manner. Neurobiol Dis 134:104666.

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci 9:357–381.

Apicella P, Deffains M, Ravel S, Legallet E (2009) Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. Eur J Neurosci 30:515–526.

Apicella P, Ljungberg T, Scarnati E, Schultz W (1991) Responses to reward in monkey dorsal and ventral striatum. Exp Brain Res 85.

Asaad WF, Eskandar EN (2011) Encoding of Both Positive and Negative Reward Prediction Errors by Neurons of the Primate Lateral Prefrontal Cortex and Caudate Nucleus. J Neurosci 31:17772–17787.

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

Banaie Boroujeni K, Oemisch M, Hassani SA, Womelsdorf T (2020) Fast spiking interneuron activity in primate striatum tracks learning of attention cues. Proc Natl Acad Sci 117:18049–18058.

Bartolo R, Prado L, Merchant H (2014) Information Processing in the Primate Basal Ganglia during Sensory-Guided and Internally Driven Rhythmic Tapping. J Neurosci 34:3910–3923.

Berke JD, Okatan M, Skurski J, Eichenbaum HB (2004) Oscillatory Entrainment of Striatal Neurons in Freely Moving Rats. Neuron 43:883–896.

Bray S, O'Doherty J (2007) Neural Coding of Reward-Prediction Error Signals During Classical Conditioning With Attractive Faces. J Neurophysiol 97:3036–3045.

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proc Natl Acad Sci 101:9849–9854.

Brovelli A, Laksiri N, Nazarian B, Meunier M, Boussaoud D (2008) Understanding the Neural Computations of Arbitrary Visuomotor Learning through fMRI and Associative Learning Theory. Cereb Cortex 18:1485–1495.

Brovelli A, Nazarian B, Meunier M, Boussaoud D (2011) Differential roles of caudate nucleus and putamen during instrumental learning. NeuroImage 57:1580–1590.

Brown P (2007) Abnormal oscillatory synchronisation in the motor system leads to impaired movement. Curr Opin Neurobiol 17:656–664.

Brown P, Williams D (2005) Basal ganglia local field potential activity: Character and functional significance in the human. Clin Neurophysiol 116:2510–2519.

Buzsáki G (2006) Rhythms of the Brain. Oxford University Press.

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420.

Calderon CB, De Loof E, Ergo K, Snoeck A, Boehler CN, Verguts T (2021) Signed Reward Prediction Errors in the Ventral Striatum Drive Episodic Memory. J Neurosci 41:1716–1726.

Cohen L (1995) Time-frequency analysis. Englewood Cliffs, N.J: Prentice Hall PTR.

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

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.
Courtemanche R, Fujii N, Graybiel AM (2003) Synchronous, Focally Modulated Beta-Band Oscillations Characterize Local Field Potential Activity in the Striatum of Awake Behaving Monkeys. J Neurosci 23:11741--11752.
Dickinson A (1980) Contemporary animal learning theory. Cambridge University Press.
Dickinson A (1994) Instrumental conditioning. Anim Learn Cogn:45–79.
Feingold J, Gibson DJ, DePasquale B, Graybiel AM (2015) Bursts of beta oscillation differentiate postperformance activity in the striatum and motor cortex of monkeys performing movement tasks. Proc Natl Acad Sci 112:13687–13692.
Fiorillo CD, Tobler PN, Schultz W (2003) Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. Science 299:1898–1902.
Fujiyama F, Takahashi S, Karube F (2015) Morphological elucidation of basal ganglia circuits contributing reward prediction. Front Neurosci 9.
Garcia S, Guarino D, Jaillet F, Jennings T, Pröpper R, Rautenberg PL, Rodgers CC, Sobolev A, Wachtler T, Yger P, Davison AP (2014) Neo: an object model for handling electrophysiology data in multiple formats. Front Neuroinformatics 8.
Goldberg JA (2004) Spike Synchronization in the Cortex-Basal Ganglia Networks of Parkinsonian Primates Reflects Global Dynamics of the Local Field Potentials. J Neurosci 24:6003–6010.
Gramfort A (2013) MEG and EEG data analysis with MNE-Python. Front Neurosci 7.
Haber SN (2003) The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat 26:317–330.
Haber SN, Fudge JL, McFarland NR (2000) Striatonigrostriatal Pathways in Primates Form an Ascending Spiral from the Shell to the Dorsolateral Striatum. J Neurosci 20:2369–2382.
Han M-J, Park C-U, Kang S, Kim B, Nikolaidis A, Milham MP, Hong SJ, Kim S-G, Baeg E (2021) Mapping functional gradients of the striatal circuit using simultaneous microelectric stimulation and ultrahigh-field fMRI in non-human primates. NeuroImage 236:118077.
Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. J Neurosci 28:5623–5630.
Howe MW, Atallah HE, McCool A, Gibson DJ, Graybiel AM (2011) Habit learning is associated with major shifts in frequencies of oscillatory activity and synchronized spike firing in striatum. Proc Natl Acad Sci 108:16801–16806.
Hoyer S, Hamman J (2017) xarray: N-D labeled Arrays and Datasets in Python. J Open Res Softw 5:10.
Hunter JD (2007) Matplotlib: A 2D Graphics Environment. Comput Sci Eng 9:90–95.
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.
Jahanshahi M, Obeso I, Rothwell JC, Obeso JA (2015) A fronto–striato–subthalamic–pallidal network for goal-directed and habitual inhibition. Nat Rev Neurosci 16:719–732.
Jenkinson N, Brown P (2011) New insights into the relationship between dopamine, beta oscillations and motor function. Trends Neurosci 34:611–618.
Keramati M, Dezfouli A, Piray P (2011) Speed/Accuracy Trade-Off between the Habitual and the Goal-Directed Processes Behrens T, ed. PLoS Comput Biol 7:e1002055.
Kilavik BE, Zaepffel M, Brovelli A, MacKay WA, Riehle A (2013) The ups and downs of beta oscillations in sensorimotor cortex. Exp Neurol 245:15–26.
Kondabolu K, Roberts EA, Bucklin M, McCarthy MM, Kopell N, Han X (2016) Striatal cholinergic interneurons generate beta and gamma oscillations in the corticostriatal circuit and produce motor deficits. Proc Natl Acad Sci 113:E3159–E3168.
Kumar P, Goer F, Murray L, Dillon DG, Beltzer ML, Cohen AL, Brooks NH, Pizzagalli DA

(2018) Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. Neuropsychopharmacology 43:1581–1588.
Leventhal DK, Gage GJ, Schmidt R, Pettibone JR, Case AC, Berke JD (2012) Basal Ganglia Beta Oscillations Accompany Cue Utilization. Neuron 73:523–536.
Marchand WR, Lee JN, Thatcher JW, Hsu EW, Rashkin E, Suchy Y, Chelune G, Starr J, Barbera SS (2008) Putamen coactivation during motor task execution. NeuroReport 19:957–960.
Marche K, Apicella P (2021) Activity of fast-spiking interneurons in the monkey striatum during reaching movements guided by external cues or by a free choice. Eur J Neurosci 53:1752–1768.
Marche K, Martel A-C, Apicella P (2017) Differences between Dorsal and Ventral Striatum in the Sensitivity of Tonically Active Neurons to Rewarding Events. Front Syst Neurosci 11.
Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.
McKinney W (2010) Data Structures for Statistical Computing in Python. Proc 9th Python Sci Conf 445:56–61.
Mestres-Missé A, Turner R, Friederici AD (2012) An anterior–posterior gradient of cognitive control within the dorsomedial striatum. NeuroImage 62:41–47.
Mitra PP, Pesaran B (1999) Analysis of Dynamic Brain Imaging Data. Biophys J 76:691–708.
Münte TF, Marco-Pallares J, Bolat S, Heldmann M, Lütjens G, Nager W, Müller-Vahl K, Krauss JK (2017) The human globus pallidus internus is sensitive to rewards – Evidence from intracerebral recordings. Brain Stimulat 10:657–663.
Nonomura S, Nishizawa K, Sakai Y, Kawaguchi Y, Kato S, Uchigashima M, Watanabe M, Yamanaka K, Enomoto K, Chiken S, Sano H, Soma S, Yoshida J, Samejima K, Ogawa M, Kobayashi K, Nambu A, Isomura Y, Kimura M (2018) Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. Neuron 99:1302-1314.e5.
O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr Opin Neurobiol 14:769–776.
O'Doherty JP (2007) Lights, Camembert, Action! The Role of Human Orbitofrontal Cortex in Encoding Stimuli, Rewards, and Choices. Ann N Y Acad Sci 1121:254–272.
O'Doherty JP, Cockburn J, Pauli WM (2017) Learning, reward, and decision making. Annu Rev Psychol 68:73–100.
O'Doherty JP, Hampton A, Kim H (2007) Model-Based fMRI and Its Application to Reward Learning and Decision Making. Ann N Y Acad Sci 1104:35–53.
Oyama K, Hernadi I, Iijima T, Tsutsui K-I (2010) Reward Prediction Error Coding in Dorsal Striatal Neurons. J Neurosci 30:11447–11457.
Panzeri S, Treves A (1996) Analytical estimates of limited sampling biases in different information measures. Netw Comput Neural Syst 7:87–107.
Parent A, Hazrati L-N (1995) Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Res Rev 20:91–127.
Park SQ, Kahnt T, Talmi D, Rieskamp J, Dolan RJ, Heekeren HR (2012) Adaptive coding of reward prediction errors is gated by striatal coupling. Proc Natl Acad Sci 109:4285–4289.
Paxinos G, Huang X-F, Petrides M, Toga A (2008) The Rhesus monkey brain : in stereotaxic coordinates, 2nd ed. San Diego, California, USA: Academic Press.
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830.
Pennartz CMA, Ito R, Verschure PFMJ, Battaglia FP, Robbins TW (2011) The hippocampal–striatal axis in learning, prediction and goal-directed behavior. Trends Neurosci 34:548–559.
Percival DB, Walden AT (1993) Spectral Analysis for Physical Applications. Cambridge University Press.

Pine A, Sadeh N, Ben-Yakov A, Dudai Y, Mendelsohn A (2018) Knowledge acquisition is governed by striatal prediction errors. Nat Commun 9.
Plotly Technologies Inc. (2015) Collaborative data science. Plotly Technol Inc Available at: https://plot.ly.
Rescorla RA (1991) Associative relations in instrumental learning: The eighteenth Bartlett memorial lecture. Q J Exp Psychol 43:1–23.
Ressler N (2004) Rewards and punishments, goal-directed behavior and consciousness. Neurosci Biobehav Rev 28:27–39.
Rezayat E, Dehaqani M-RA, Clark K, Bahmani Z, Moore T, Noudoost B (2021) Frontotemporal coordination predicts working memory performance and its local neural signatures. Nat Commun 12:1103.
Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G (2009) Ventral Striatal Neurons Encode the Value of the Chosen Action in Rats Deciding between Differently Delayed or Sized Rewards. J Neurosci 29:13365–13376.
Salazar RF, Dotson NM, Bressler SL, Gray CM (2012) Content-Specific Fronto-Parietal Synchronization During Visual Working Memory. Science 338:1097–1100.
Schmidt R, Leventhal DK, Mallet N, Chen F, Berke JD (2013) Canceling actions involves a race between basal ganglia pathways. Nat Neurosci 16:1118–1124.
Schultz W (2006) Behavioral Theories and the Neurophysiology of Reward. Annu Rev Psychol 57:87–115.
Schultz W (2007) Multiple Dopamine Functions at Different Time Courses. Annu Rev Neurosci 30:259–288.
Schultz W (2016a) Dopamine reward prediction-error signalling: a two-component response. Nat Rev Neurosci 17:183–195.
Schultz W (2016b) Dopamine reward prediction error coding. Dialogues Clin Neurosci 18:10.
Schultz W (2016c) Reward functions of the basal ganglia. J Neural Transm 123:679–693.
Schwerdt HN, Amemori K, Gibson DJ, Stanwicks LL, Yoshida T, Bichot NP, Amemori S, Desimone R, Langer R, Cima MJ, Graybiel AM (2020) Dopamine and beta-band oscillations differentially link to striatal value and motor control. Sci Adv 6:eabb9226.
Seabold S, Perktold J (2010) Statsmodels: Econometric and Statistical Modeling with Python. Proc 9th Python Sci Conf 57:10–25080.
Silvetti M, Nuñez Castellar E, Roger C, Verguts T (2014) Reward expectation and prediction error in human medial frontal cortex: An EEG study. NeuroImage 84:376–382.
Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2012) Reward Prediction Error Signaling in Posterior Dorsomedial Striatum Is Action Specific. J Neurosci 32:10296–10305.
Sutton RS, Barto AG (1998) Reinforcement learning: an introduction, Cambridge, MA: The MIT Press.
Valentin VV, O'Doherty JP (2009) Overlapping Prediction Errors in Dorsal Striatum During Instrumental Learning With Juice and Money Reward in the Human Brain. J Neurophysiol 102:3384–3391.
Vezoli J, Vinck M, Bosman CA, Bastos AM, Lewis CM, Kennedy H, Fries P (2021) Brain rhythms define distinct interaction networks with differential dependence on anatomy. Neuron 109:3862-3878.e5.
Vogelsang DA, D'Esposito M (2018) Is There Evidence for a Rostral-Caudal Gradient in Fronto-Striatal Loops and What Role Does Dopamine Play? Front Neurosci 12.
Wasserman EA, Miller RR (1997) What's elementary about associative learning? Annu Rev Psychol 48:573–607.
Watkins CJ, Dayan P (1992) Q-learning. Mach Learn 8:279–292.