BETA-BAND POWER PREDICTS EXPLORATORY CHOICES IN PROBABILISTIC ENVIRONMENT

*A.S. Miasnikova, V.D. Tretyakova, K.E. Sayfulina, G.L. Kozunova, A.M. Rytikova, B.V. Chernyshev**

Moscow State University of Psychology and Education, 29 Sretenka str., Moscow, 127051, Russia

* Corresponding author: chernyshevbv@mgppu.ru

Abstract. In probabilistic conditions, people choose low-payoff alternatives on some trials, thus failing to maximize their payoffs. We suggest that such behavior implicates exploration of task rules by choosing risky options instead of exploiting more rewarding alternatives. We hypothesized that exploration would affect brain responses to feedback. Further, a shift to exploration develops gradually and, therefore, a decision to make an exploratory choice may be observed on trials preceding risky choices. We investigated beta power (16–30 Hz) in the magnetoencephalographic data from 62 healthy participants performing a two-choice probabilistic gambling with monetary gains and losses. The effects were found at 600–800 ms after feedback onset in frontal, central and occipital brain regions. On trials preceding risky choices we identified a decrease in beta power which implies a change in decision-making strategy and a shift towards cognitive flexibility and exploration. An increase in beta power during risky decisions indicates that reward learning mechanisms are implicated. Increases in beta power following losses in risky choices indicates at the process of updating the internal representation of the task. In summary, current findings reveal that the outcomes of exploratory trials are processed differentially, while there is no evidence of such processing on exploitatory trials. This corroborates the hypothesis that exploratory choices represent active probing into the surmised task rules. Current findings also suggest that the processing of outcomes preceding the exploratory trials is altered in such a way that subjects override their intention to use the utility model and reset their behavioral strategy.

Keywords: decision-making, gambling, probabilistic outcome, exploration, magnetoencephalography.

List of abbreviations

MEG – magnetoencephalography

Introduction

In a probabilistic environment people often tend to choose low-payoff options, thus failing to maximize their gains (Shanks *et al.*, 2002). One explanation of such behavior is that humans develop a set of expectations about the task regularities, and test their beliefs by exploring risky low-payoff alternatives instead of consistently exploiting the rewarding options (Sayfulina *et al.*, 2020; Cogliati Dezza *et al.*, 2017). This behavior may seem to be in striking contradiction to reward learning and rational choice, although such a strategy may be adaptive in the long run in realistic situations (Friston *et al.*, 2015; Parr & Friston, 2017). Neuronal and cognitive processes underlying internally triggered shifts to exploratory behavior are largely unknown.

In the current study, we aimed to reveal neurophysiological correlates of a change in decision-making towards exploratory strategy in the 2-choice probabilistic gambling task. Based on the evidence that risky choices of low-payoff options could be acts of intentional probing into surmised task rules (Sayfulina *et al.*, 2020; Cogliati Dezza *et al.*, 2017), we hypothesized that, first, greater importance of outcomes on risky exploratory trials compared with exploitatory ones will induce differential neuronal sensitivity to the positive versus negative feedback regarding the outcome of the risky choice. Second, we hypothesized that a decision to switch to exploration of task rules is not accidental or instantaneous (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Jepma & Nieuwenhuis, 2011), and some covert preparation to make a change in the strategy towards exploration might take place before the exploratory choices themselves: therefore, we expected that some correlates of readiness to take risks could be detected on trials preceding exploratory choices. Based on these assumptions, we expected to observe altered brain responses to feedback on exploitatory trials preceding risky exploratory choices.

We focused our analysis on oscillations in the beta band because this frequency range is related to reward processing (HajiHosseini *et al.*, 2012; Luft, 2014) and is indicative of keeping or changing cognitive sets (Engel & Fries, 2010).

Methods

Participants. 62 healthy participants (31 male, 34 female, mean age = 23, SD = 6.5) having no records of neurological or severe visual impairment took part in the study.

Ethics statement. This study was carried in accordance with the Declaration of Helsinki; the study protocol was approved by the Ethics Committee of Moscow State University of Psychology and Education. All participants gave written informed consent.

Experimental paradigm. We used a modified version of a probabilistic gambling task. Participants were instructed to select one of the two figures presented on the screen simultaneously by pressing a corresponding button (Fig. 1). Each pair of figures was derived from a hiragana hieroglyph rotated at different angles. A new pair of figures was presented in each experimental block. Selection of one figure was followed by a gain more often than by a loss (70%/30%, high-payoff option), while the probabilities were reversed for the other figure in each pair (30%/70%, low-payoff option). A positive or negative feedback (a gain or a loss) was delivered 1000 ms after pressing the button. Participants were supposed to learn which figure was rewarded more often and press the corresponding button. Upon completing each block, participants were presented with their cumulative score.

Trial selection. Within each experimental block, we considered only trials that followed reaching the learning criterion (four consecutive choices of a high-payoff stimulus). Selection of a low-payoff stimulus during these periods were considered to be exploratory choices (“risk”). We also analyzed exploitatory trials immediately preceding exploratory choices (“prerisk” correspondingly). Reference condition was exploitatory choices not neighboring exploratory choices (“norisk”).

MEG data processing. MEG data were recorded using a 306-channel MEG system (Elekta Neuromag VectorView). The analysis of MEG data was performed with custom scripts using MNE-python software. Independent component analysis (ICA) was used to remove biological artefacts. Power in the beta frequency range (~16–30 Hz) was calculated using multitaper time-frequency analysis, for each trial separately.

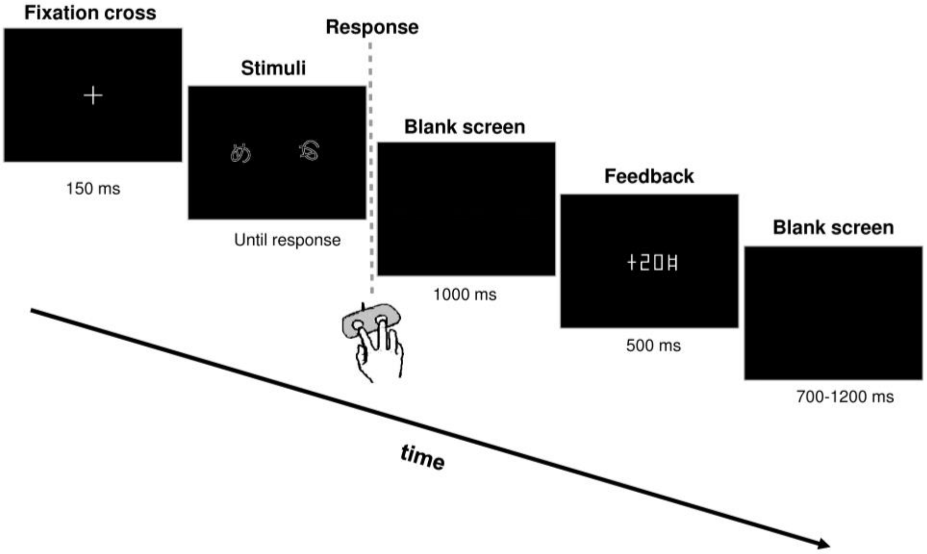

Fig. 1. Experimental paradigm (see text for details)

Statistics. Our analysis aimed to investigate two main contrasts: (1) exploratory trials ("risk") vs. exploitatory trials ("norisk"), and (2) exploitatory trials preceding exploration ("prerisk") vs. exploitatory trials ("norisk") aiming at pinpointing neural mechanisms underlying decision-making prior to and during making an exploratory ("risky") choice. The area and time of interest were obtained by applying a linear mixed model (LMM) to all combined planar gradientometers for the interaction of choice type ('norisk', 'prerisk', 'risk', 'postrisk') x feedback sign (negative and positive) with False Discovery Rate (FDR) correction for multiple comparisons. The interaction was used in accordance with the expected differential responses to feedback sign depending upon choice types. Specifically, we expected higher feedback salience in exploratory choices. We identified significant channels in the time interval as of 600–800 ms post-feedback and considered this time interval in further statistical comparisons. Channels with p-value < .05 were following FDR correction for multiple comparisons were taken for the further analyses. For further statistical comparisons we used linear mixed-effects model (LMMs) with the following fixed effects: 'choice type' (4 levels: 'norisk', 'prerisk', 'risk' and 'postrisk'), 'feedback sign' (two levels: 'gain' and 'loss') and their interaction, subjects as the random effect, and beta power averaged in 600–800 ms post-feedback as the dependent variable. The Tukey HSD test for multiple comparisons was used for post hoc analyses. The results are reported for p-values < 0.05.

Results

We have revealed 37 significant combined planar gradientometers located over frontal, central and occipital areas at 600–800 ms after feedback onset (*p* < .05, FDR-corrected (Fig. 2A). We report the following results for those gradientometers.

Fig. 3 illustrates the topographical maps of beta power across choice types: in exploratory ('risk') trials (Fig. 3A), in trials preceding exploratory ones ('prerisk') (Fig. 3B), in exploitatory ('norisk') trials (Fig. 3C).

LMM analysis of post-feedback beta revealed a significant effect of 'choice type' (F3, 4320 = 9.48, *p* < .001). Post hoc analysis showed that the beta power on "prerisk" trials was significantly smaller compared to the other

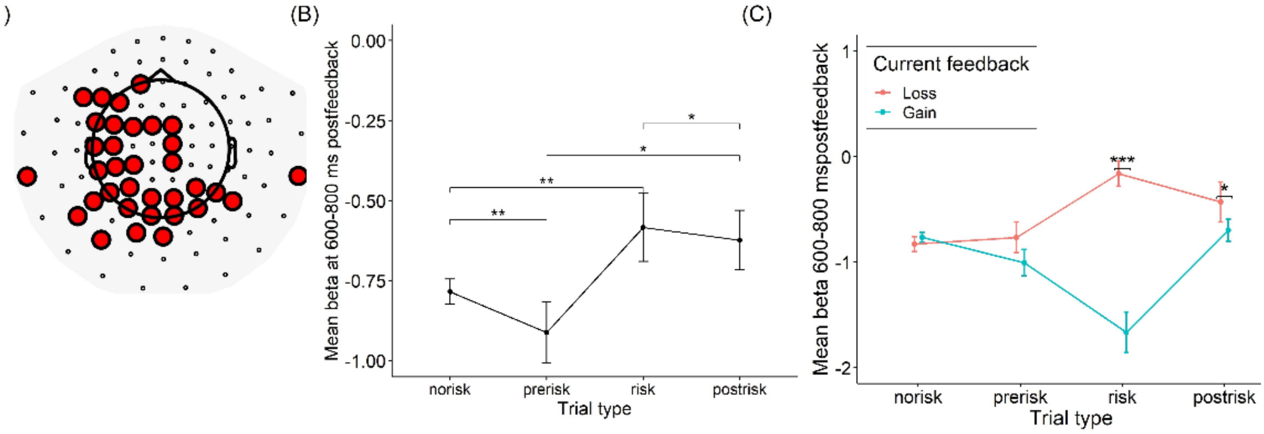

Fig. 2. Feedback-related parietal beta power differentiates exploratory "risk" trials from exploitatory choices.
A – significant sensors derived from the fixed effects ANOVA interaction (choice type x feedback sign) at 800 ms after feedback onset (*p* < .05, FDR-corrected); B – beta power averaged over significant sensors as a function of choice type; C – beta power averaged over significant sensors split by feedback sign in each choice type. Post hoc comparisons were performed using Tukey's test. In (B) and (C), means and standard errors are shown.
*** *p* < .001, * *p* < .05, **.001 < *p* < .01.

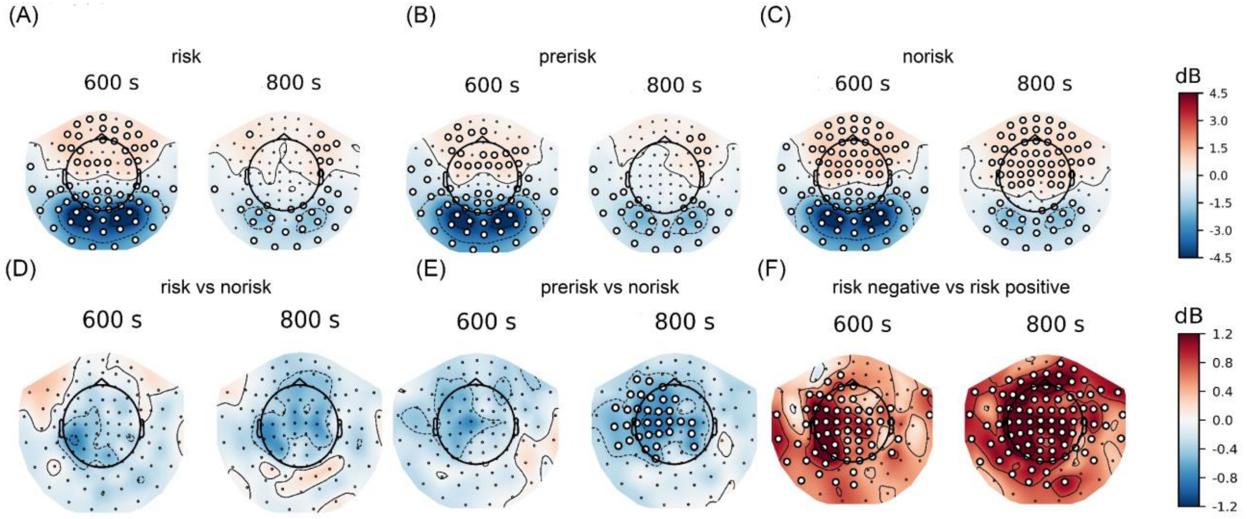

Fig. 3. Feedback-related beta power differentiates exploratory "risk" trials from exploitatory types at 600-800 ms after feedback onset. A – topographical map of beta power in exploratory "risk" choice type; B – topographical map of beta power in "prerisk" choice type; C – topographical map of beta power in exploitatory "norisk" choice type; D – topographical map of difference in beta power between "risk" and "norisk" choice types. E – topographical map of difference in beta power between "prerisk" and "norisk" choice types; F – topographical map of difference in beta power between negative and positive feedback signs on the "risk" choice type. Statistically significant sensors derived from t-tests are marked by white circles (p < .05, FDR-corrected)

exploitatory choice types ("norisk" and "postrisk"), whereas beta in risky decisions was the highest (Fig. 2B, 3D, 3E). We also found a significant effect of 'feedback sign' (F1, 4317 = = 33.48, p < .001) and a significant interaction of 'choice type' and 'feedback sign' (F3, 4318 = 13.78, p < .001). Post hoc comparisons involving feedback sign showed that the beta power was significantly greater after losses compared with gains on "risk" and "postrisk" trials only (Fig. 2C, 3F).

Discussion

In our everyday live, we have to make choices, and many choice options imply uncertainty or risk. If people behaved rationally, people would prefer options with the highest expected value, but, in practice, different choices may prevail in an attempt to explore the option with the potential to obtain the highest possible outcome in future. Here, we investigated an internally triggered shift from safe exploitatory behavior to risky exploratory one using a probabilistic binary choice task with a constant pay-off probability. Exploratory behavior manifests itself as rare spontaneous choices of a low-payoff risky option made by the participants, who by trial-and-error learning acquired knowledge of expected values and demonstrated stable preference towards the high-payoff option, i.e. exploitatory behavior. Such exploratory choices might represent intentional behavioral acts committed to maximize the benefit at the expense of the current utility model.

Using MEG beta power as a neural signature of maintaining or changing a current cognitive set (Spitzer & Haegens, 2017), we found that risky exploratory choices differed from the exploitatory behavior in the way how the brain processed the feedback signal about the choice outcome. Risky exploratory choices, in contrast to exploitatory ones, were followed by dramatically increased differential sensitivity of beta power to the feedback sign. While reward for risky choice was accompanied by a highly significant drop in the beta power, its punishment was followed by an equally significant beta increase. This differential effect was most pronounced over the frontal, central and occipital brain areas. The localization and timing of the effect comply with those reported by Yaple *et al.* (2018).

Strong differences in the neural beta responses to losses and gains after exploratory (risky) choices support the idea that exploratory choices represent active probing into the surmised task rules (Sayfulina *et al.*, 2020; Cogliati Dezza *et al.*, 2017), and the outcomes of such choices may imply learning task regularities. Since rewarding feedback during exploratory trials contradicts the utility model acquired by participants and thus remains unexpected, it may trigger model updating (HajiHosseini *et al.*, 2012). This interpretation fits well with the supposed role of beta power suppression during the changes of cognitive set (Spitzer & Haegens, 2017). Increases in beta power following gains are described in previous literature and interpreted in terms of a specific role of beta oscillations in the tendency to adhere to the current motor/cognitive set (Luft, 2014).

Unlike discriminative beta reactivity to losses and gains on risky trials, beta power was strongly suppressed after both positive and negative feedback signals on trials that preceded risky choices. In the view of the functional role of beta suppression, this suggests that processing of choice outcomes is altered in such a way that subjects override their intention to use the utility model and intend to reset their behavioral program regardless of the outcome of their current act. The involvement of left central areas in the effect might imply a change in decision-making towards a more flexible strategy (Luft, 2014), whereas frontal areas suggest reward learning processes. Considering the two findings combined, we suggest that we have observed a gradually evolving shift towards exploration. Although the effects observed in the current study were phasic, the finding of the effects that covertly precede the exploratory choice, are compatible with the notion of tonic changes in noradrenergic efflux being an important part of the mechanism that predisposes the subject towards explorations (Aston-Jones & Cohen, 2005; Usher *et al.*, 1999; Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011).

Our findings suggest that the neurocognitive mechanism of a shift towards exploration may imply at least two stages: (1) decreased salience of outcomes on a trial preceding exploration, suggesting readiness to abandon the current decision-making strategy and reset the current behavioral program, and (2) increased processing of outcomes of exploratory choices, which may induce utility model updating.

Acknowledgements

Supported by Russian Science Foundation grant 20-18-00252.

Authors declare no conflicts of interest.

References

ASTON-JONES G. & COHEN J.D. (2005): An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience* **28**, 403–450.

COGLIATI DEZZA I., YU A.J., CLEERMANS A. & WILLIAM A. (2017): Learning the value of information and reward over time when solving exploration-exploitation problems. *Scientific Reports* **7**, 16919. doi: 10.1038/s41598-017-17237-w.

ENGEL A.K. & FRIES P. (2010): Beta-band oscillations - signalling the status quo? *Current opinion in neurobiology* **20**(2), 156-165. doi: 10.1016/j.conb.2010.02.015.

PARR T. & FRISTON K.J. (2017): Uncertainty, epistemics and active inference. *Journal of The Royal Society Interface* **14**(136), 20170376.

FRISTON K., RIGOLI F., OGNIBENE D., MATHYS C., FITZGERALD T. & PEZZULO G. (2015): Active inference and epistemic value. *Cognitive Neuroscience* **6**(4), 187–214. doi: 10.1080/17588928.2015.1020053.

GILZENRAT M.S., NIEUWENHUIS S., JEPMA M. & COHEN J.D. (2010): Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive Affective & Behavioral Neuroscience* **10**(2), 252-269. doi: 10.3758/Cabn.10.2.252.

JEPMA M., BEEK E.T.T., WAGENMAKERS E.J., VAN GERVEN J.M.A. & NIEUWENHUIS S. (2010): The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. *Frontiers in Human Neuroscience* **4**. doi: 10.3389/Fnhum.2010.00170.

HAJIHOSSEINI A., RODRIGEZ-FORNELLS A. & MARCO-PALLARES J. (2012): The role of beta-gamma oscillations in unexpected rewards processing. *NeuroImage* **60**(3), 1678-85. doi: 10.1016/j.neuroimage.2012.01.125.

LUFT C.D.B. (2014): Learning from feedback: the neural mechanisms of feedback processing facilitating better performance. *Behavioral Brain Research* **261**, 356–368. doi: 10.1016/j.bbr.2013.12.043.

SAYFULINA K.E., KOZUNOVA G.L., MEDVEDEV V.A., RYTIKOVA A.M. & CHERNYSHEV B.V. (2020): Decision making under uncertainty: exploration and exploitation [Elektronnyi resurs]. *Sovremennaia zarubezhnaia psikhologiia = Journal of Modern Foreign Psychology* **9**(2), 93–106. doi:10.17759/jmfp.2020090208 (In Russ., abstr. in Engl.).

SHANKS D.R., TUNNEY R.J. & MCCARTHY J.D. (2002): A re-examination of probability matching and rational choice. *J Behav Decis Mak* **15**(3), 233–250. doi: 10.1002/bdm.413.

SPITZER B. & HAEGENS S. (2017): Beyond the status quo: a role for beta oscillations in endogenous content (re) activation. *eNeuro*, **4**(4).

USHER M., COHEN J.D., SERVAN-SCHREIBER D., RAJKOWSKI J. & ASTON-JONES G. (1999): The role of locus coeruleus in the regulation of cognitive performance. *Science* **283**(5401), 549–554.

YAPLE Z., MARTINEZ-SAITO M., NOVIKOV N., ALTUKHOV D., SHESTAKOVA A. & KLUCHAREV V. (2018): Power of Feedback-Induced Beta Oscillations Reflect Omission of Rewards: Evidence from an EEG Gambling Study. *Frontiers in Neuroscience* **12**, 776. doi: 10.3389/fnins.2018.00776.