**Reply to Reviewers**

In this reply, Reviewer and Editor comments are in black and our replies in blue. *Italic font style* is used wherever text from the revised manuscript is quoted.

**Editor**

Your revisions should address the specific points made by each reviewer. Please note that reviewer 4 still has lingering concerns with the use of 'mutual learning'. You would need to significantly improve your rationale to justify why you are using this term. Please submit a point-by-point response to all of the reviewers' comments that indicates the changes you have made to the manuscript. This should be uploaded to the ‘Response to Reviewers’ card in Aperta alongside a 'track-changes' version of your manuscript that specifies the edits made.

Dear Editor of PLoS Biology,

We would like to thank you for positively evaluating our revised manuscript and considering it for publication in PLOS Biology. In this document, we provide a point-by-point reply to all remaining concerns raised by the reviewers and refer to the corresponding amendments in our further revised manuscript.

We are confident that these additional modifications and responses address all remaining issues and hope for the final acceptance of our manuscript.

Sincerely yours,

The authors

**Reviewer 1**

Comment 1.1: “Reviewer 1: No comments to authors”

We sincerely thank this reviewer for a very constructive criticism on our original submission which has greatly improved our manuscript, especially helping us to refocus its content and clarify the hypothesis.

**Reviewer 2**

Comment 2.1: “This manuscript has been markedly improved by the revisions. I have two recommendations for further changes.”

We would like to thank this reviewer for his considerable contribution to the improvement of our manuscript, especially with respect to the discussion of our findings against the relevant literature, as well as for appreciating our revision. We make every effort in this response letter to address the two remaining concerns, providing additional analyses and justifications.

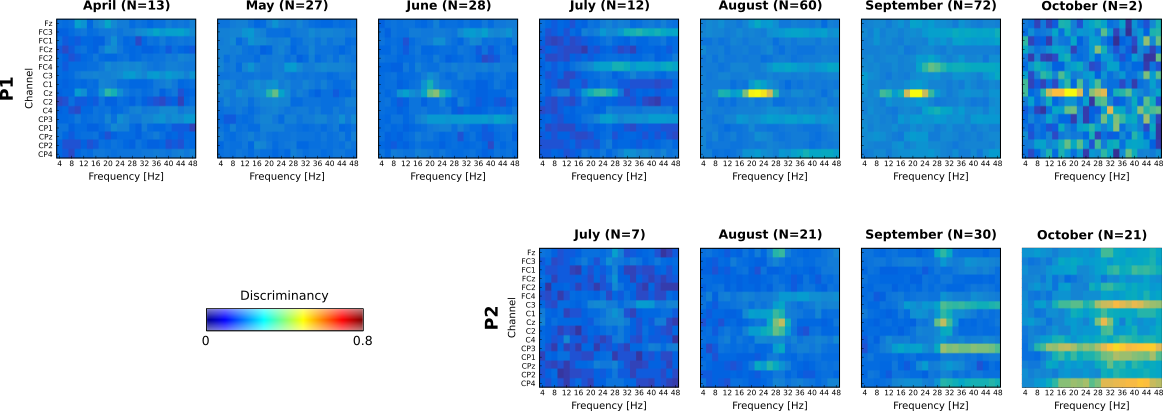
Comment 2.2: “First, the figure showing the frequency specificity of the BCI control developed by the subjects should be included in the manuscript. It is as important as the multiple topographies that are now included; it should not be relegated to the supplementary figures. Also, while the frequency specificity is quite clear for P1, P2 seems to show higher-frequency activity later on. Thus, it would be preferable if the figure included higher frequencies for P2 to show the full frequency distribution of that control. Its clear topographic specificity implies that it is not EMG activity. However, if the control extends well above 30-40 Hz, the authors will need to discuss the issue.”

Thank you for giving us the opportunity to clarify these important points.

We agree with the reviewer that the spectral distribution of SMR patterns is equally important to the spatial distribution. In fact, Figure 4A provides both components because, for both subjects (P1 and P2), the spectral components that evolved over training where located in the beta band only.

The additional information in the discriminancy maps (Fig. S1) is: a) the specific beta sub-bands activated for each user and b) demonstration of the claim made in the text that all spectral content is concentrated in the beta band. We think this extra information is not sufficiently important to justify the inclusion of both figures in the main text or the replacement of the topographic maps (which show spatial information more effectively) by the spectral ones. The reader can always refer to the supplementary material for these details. We have opted to maintain the topographic maps in the text body and the spectral maps in the supplementary material. **Clearly, should the reviewer and editor remain unconvinced by the reasoning put forward here, we are willing to move both figures in the text body for the final version of the manuscript.**

Regarding the observation that the lateral SMR patterns of user P2 might be affected by some kind of high-frequency artifact, indeed, the reviewer is right. Following the reviewer’s recommendation, we have closely examined the higher frequencies and extended the illustrated spectral content shown in Fig. S1 to 48 Hz (just below the notch filter suitable for the European powerline noise of 50 Hz). In this figure (Fig. R1 in this reply), it can be clearly seen that 3 lateral channels (among which the only selected channel is CP3) exhibit an unidentified high-frequency component in September and, especially, in October. We agree with the reviewer that this cannot be EMG activity, but we have no reasonable assumption regarding its origin.



**Figure R1. BCI feature discriminancy maps per training month.**

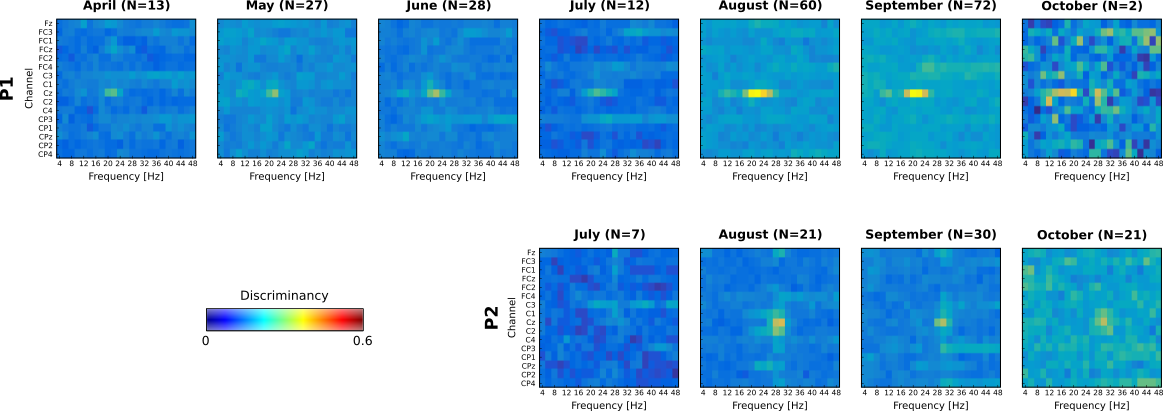
However, it must be underlined that this effect is not particularly alarming regarding the integrity of our data, for the following reasons:

a) Inspection of the discriminancy maps of each individual run that took place within these months shows that only a minority of the total runs (20 out of 79) were prominently affected by such a component in the selected channels, which is however strong enough in these few cases to survive on the monthly averages shown in the figure.

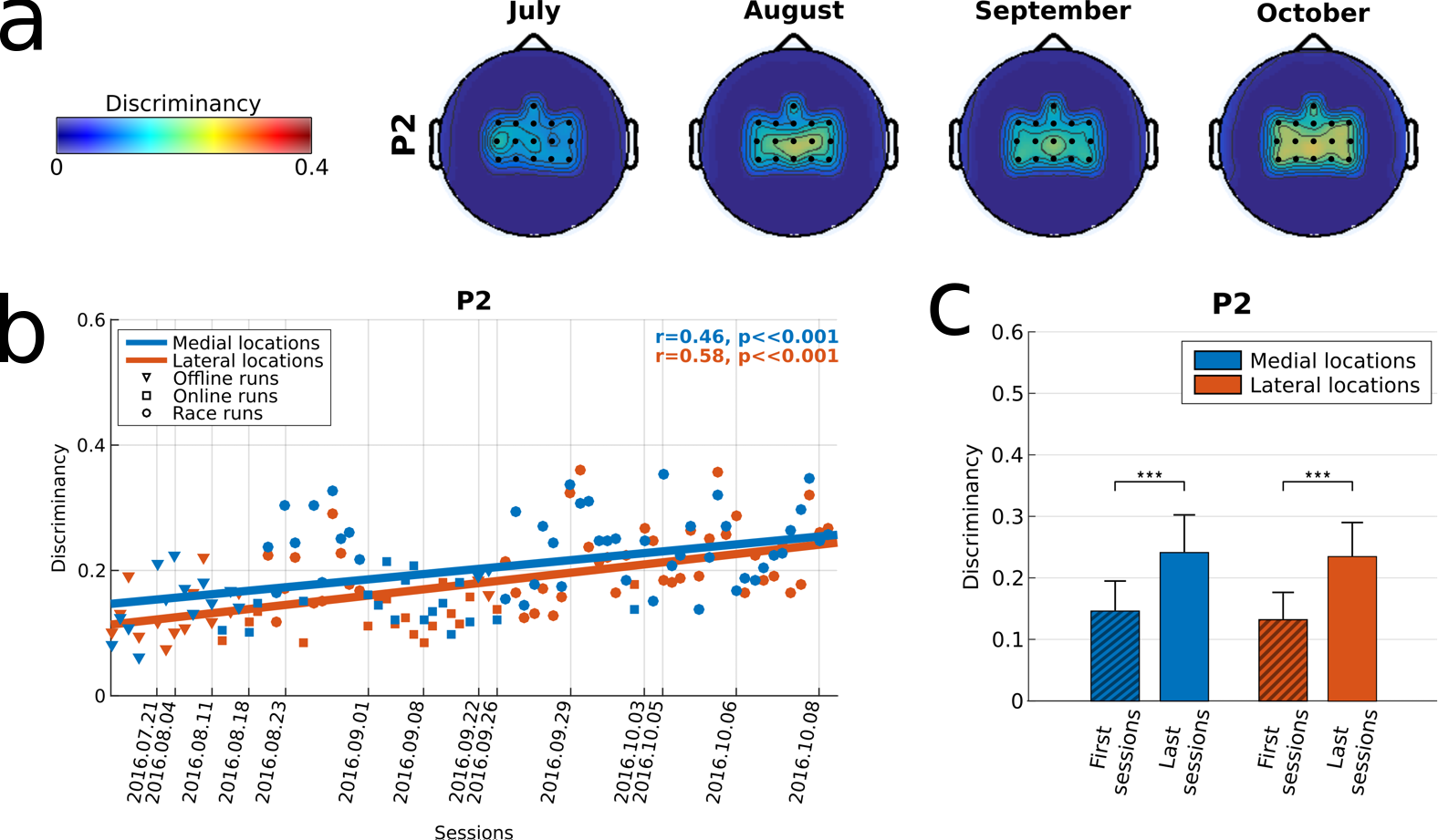
b) This undesired effect concerns only the lateral pattern of subject P2: there is absolutely no doubt regarding the intergrity of the characterization of P1’s SMR modulation and the medial emerging pattern of P2 (channel Cz), so that our main claim of demonstrating the existence and extent of subject learning for both users is well supported in any case.

c) Most importantly, only features CP3-30 Hz and CP3-32 Hz have been selected for control (Table 2 in the manuscript). Any eventual noise on CP3 above these frequencies could not have influenced P2’s control of the BCI.

There still exists some ambiguity regarding whether the discriminancy of these two selected features has a brain origin correlated to a MI task. We provide below evidence that this is in fact the case. Specifically, Fig. R2 in this reply shows the same discriminancy maps for both users after applying the popular, state-of-the-art artifact removal algorithm “FORCe” (Daly et al., 2015). Evidently, the suspicious component is almost elminated. Critically, the previously found emerging discriminant patterns for both users survive the artifact removal, confirming that they represent EEG SMR modulation. Fig. R3 in this response letter better substantiates this claim for P2 (panels for P1 are skipped here in the interest of space but can be provided upon request), showing also that the correlations of discriminancy (including the one on lateral locations) with time remain virtually unaffected in both magnitude and significance. **Literally all effects shown in the manuscript regarding discriminancy (notably, also the correlations of discriminancy with race time, accuracy and time-on-pad) hold for both users, after artifact removal with FORCe.**

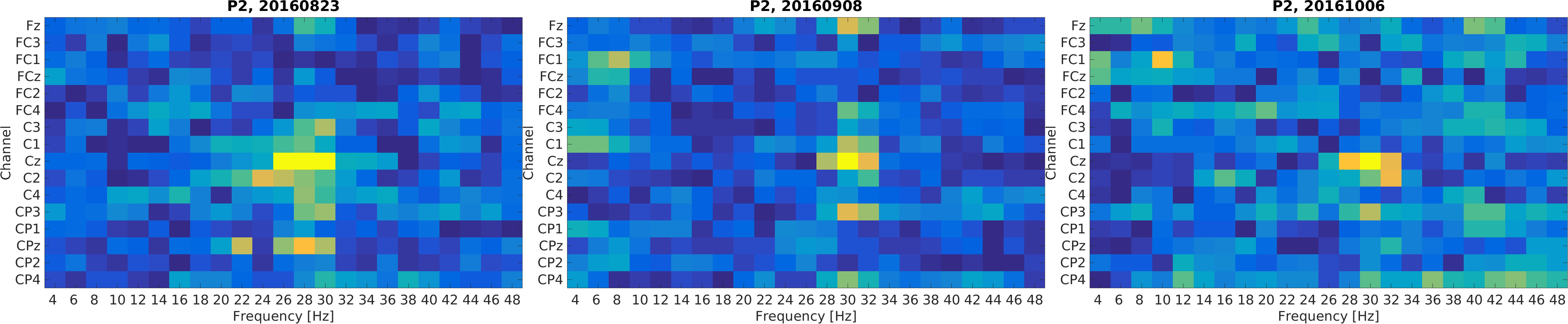


**Figure R2. BCI feature discriminancy maps per training month. Artifact removal with the FORCe algorithm has been applied on the data.**

**Figure R3. BCI feature discriminancy after artifact removal with FORCe.** **(a)** Topographic maps of discriminancy per training month for pilot P2 **(b)** Average medial (blue, channels: FCz, Cz, CPz) and lateral (red, channels FC3 ,C3, CP3, FC4, C4, CP4) discriminancy for all performed offline, online and racing runs of pilot P2. The corresponding linear fits and Pearson correlation coefficients (significance tested with Student’s t-distribution) are reported to indicate training effects. Vertical dashed lines indicate the training session where each run took place. **(C)** Average and standard deviations of medial region (blue) and lateral region (red) discriminancy within the first and last four runs of training for pilot P2. Statistically significant differences are shown with two-sided Wilcoxon ranksum tests (\*\*\*): p<.001.

Additonally, the map of August in Fig. R2 shows that the two features in question are still fairly discriminant in the absence of any high frequency component, while in September, despite some minimal high-frequency component survives the artifact removal, these two features are distinctly more discriminant (Fisher Score around 0.3) than the adjacent higher-frequency component. We thus believe that this discriminancy is of physiological origin and simply superimpsoed to unidentifiable noise sources starting around this frequency range and extending to higher frequencies. This claim is further substantiated by inspecting individual session discriminancy maps, where CP3-30 Hz and CP3-32 Hz are disriminant in the absence (or only minimal presence) of any such potential noise. Three such typical examples from August through October are shown in Fig. R4 of this reply, below, in order to clear up any doubt.

We have opted to maintain in the revised manuscript the original results related to discriminancy (without FORCe) despite, as said, all effects are preserved after FORCe. The reason is that an eventual artifact contamination has affected only a minor part of the results (maximum 3 lateral channels of P2, out of which only one was selected for control) and this influence has been shown here to be miminal. We prefer to report in the manuscript the “real” data. We find always preferable to apply only the absolutely necessary processing on the raw data, especially since one distinct feature of our study is BCI operation in real-world conditions, where such issues are common. Only the new Fig. S1 (Fig. R2 in this reply) shows data after FORCe, as acknowledged in the new caption. **Again, should the reviewer think otherwise, we are willing to proceed with the replacement of all discriminancy-related figures for the final version of the manuscript (in which case, we would prefer to do it for both users).**



**Figure R4. Typical race sessions of P2 with discriminant CP3 features at 30 and 32 Hz in August, September and October.**

Daly I, Scherer R, Billinger M, Müller-Putz G. FORCe: Fully Online and automated artifact Removal for brain-Computer interfacing. IEEE Trans Neural Syst Rehabil Eng, 23(5), 725-36, 2015.

Comment 2.2: “Second, in the Discussion paragraph that begins “The most complete evidences of…” the statements about reference 42 should be corrected. First, that paper reports results from four users, not just one. Second, Figure 2 of that paper clearly shows how individual BCI users improved with training. The statement that “across users, performance did not correlate with amount of training” simply means that some users learned faster and better than others; a person with more total training sessions was not necessarily better than a person with fewer total sessions. Nevertheless, as Figure 2 shows, each individual user improved steadily over sessions”

We sincerely apologize for this misunderstanding. This paragraph has been modified in the further revised manuscript to read *“The most complete evidence of subject learning with obvious translational implications is offered in [9], [10] and [42]. These works report on longitudinal training and involve end-users. Furthermore, [10] and [42] substantiate learning effects with ERD/ERS maps and SMR topographies, respectively, over 3-4 time points throughout the training period. Nevertheless, these works do not explicitly relate induced brain rhythm changes to BCI performance or show that SMR improvements were consistent and continuous.”*

**Reviewer 3**

Comment 3.1: “Reviewer 3 declined to re-review”

We thank this reviewer for contributing to the improvement of our manuscript in the first review round.

**Reviewer 4**

Comment 4.1: “The manuscript has been improved tremendously, it now reads much better for a neuroscience journal. I very much appreciate the authors responses to my comments. Pretty much all of my comments are now addressed adequately, but i am still not quite convinced of the term ‘mutual learning’, since for me the study essentially shows the benefit of updating classifiers while the pilot is learning. Mutual for me means that both systems learn from each other.”

We would like to thank the reviewer for acknowledging the merits of our revision.

If we interpret correctly, the reviewer considers that there is a fundamental difference between two agents that are learning in parallel and two agents that learn from each other –only the latter is mutual learning. While we agree with this distinction, we also believe that our study falls into the mutual learning category because the subjects learn in a closed loop, based on feedback provided by the machine, and the classifier is updated with new training examples provided by the subject. We suspect the reviewer might consider that only online, and not periodical as we do, classifier adaptation qualifies as mutual learning. However, as discussed in the paper, continuous online adaptation might even have detrimental effects on subject learning. Thus, mutual learning does not necessarily require the two learning agents to adapt at the same “rate”.

Therefore, we have opted to maintain this terminology in our further revised manuscript. In case the reviewer remains sceptical about the suitability of this term, we are always willing to replace it with a term that the reviewer proposes, as long as the new term still conveys the message that “both agents learn from each other”.

Comment 4.2: “just a comment on the authors reply: i am not concerned that motivation is a negative aspect. But the intense training by the pilots to get to the winning performance point begs the question whether this type of BCI is feasible for other users who may not be so externally encouraged.”

Thank you for this comment. We agree on the critical role of motivation for successful acquisition of BCI skills. Clearly, one can hardly organize an international competition in order to train subjetcs in a mutual learning approach. However, we believe novel training protocols can take this issue into account and effectively optimize the motivational aspect in various other ways: game-like designs, BCI training “in groups” so as to add the element of “competition”, immersive training environments, etc. We thus do not see motivation as a prerequisite that limits the applicability of the proposed protocol but, rather, as an influencial factor that can easily be accommodated so as to maximize the shown effects.

We have added this sentence to the discussion: *“Based on this experience, we believe that novel motivational paradigms should consider incorporating the element of “competition”, for instance, training with multi-player games.”*

Comment 4.3: “Concerning the discussion on a graded training scheme and the training effects on performance it seems logical to include a paper by Vansteensel ea on intracranial motor BCI which describes both issues and shows that even with cortical surface electrodes carefully managed training is necessary (Vansteensel et al, New England Journal of Medicine 2016)”

Thank you for this observation. This reference has been added in the Discussion of the revised manuscript, where incremental learning is discussed: *“...another contributing factor to successful SMR enhancement might have been that we have implemented an “incremental learning” approach as advocated in [53] and shown in [69], where...”*

Comment 4.4: “suggestions:

change bi-weekly to twice a week since it can also mean every 2 weeks

‘evidences’ is not really a noun, please change to ‘evidence’

fig5 needs clarification, perhaps an arrow to the selected electrodes can help

line 418 once BCI command accuracy saturated to high levels (Fig 3A), it continuous to increase for both subjects .. [should be continues]

line 350 This was also reported by our pilot P2 [44]. However, such claims are rather qualitative to qualify as hard .. [what does ‘qualitative to qualify’ mean?]”

Thank you for this suggestions. All these issues have been resolved in the revised manuscript.