



**HAL**  
open science

## Which user's state and trait can benefit from biased feedback in motor imagery BCI?

Jelena Mladenović, Jeremy Frey, Smeety Pramij, Jeremie Mattout, Fabien Lotte

### ► To cite this version:

Jelena Mladenović, Jeremy Frey, Smeety Pramij, Jeremie Mattout, Fabien Lotte. Which user's state and trait can benefit from biased feedback in motor imagery BCI?. 2021. hal-03233170v1

**HAL Id: hal-03233170**

**<https://inria.hal.science/hal-03233170v1>**

Preprint submitted on 23 May 2021 (v1), last revised 22 Dec 2021 (v2)

**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.

# Which user's state and trait can benefit from biased feedback in motor imagery BCI ?

Jelena Mladenović<sup>1,3</sup>, Jeremy Frey<sup>2</sup>, Smeety Pramij<sup>1</sup>, Jeremie Mattout<sup>3</sup>, and Fabien Lotte<sup>1</sup>

<sup>1</sup>Inria Bordeaux Sud-Ouest, LaBRI (CNRS, Univ. Bordeaux, INP), France, <sup>2</sup>Ullo, La Rochelle, France, <sup>3</sup>Lyon Neuroscience Research Centre, COPHY team, (INSERM UMRS 1028, CNRS UMR 5292, Univ. Claude Bernard Lyon 1, F-69000) Lyon, France,

**Abstract—Objective.** To control a brain-computer interface (BCI) requires its user to attain neural self-regulation. This can be arduous, especially in motor imagery BCIs. Various training methods were proposed to assist users in accomplishing BCI control and increase performance, notably the use of biased feedback, i.e. motivating (non-realistic) representation of performance. However, benefits of biased feedback on performance and learning vary between users (e.g. depending on their initial level of BCI control) and remain speculative. To disentangle the speculations about what personality type, initial state and calibration performance (CP) could benefit from a biased feedback, we conduct an experiment (n=30 for 2 sessions). **Methods.** The feedback provided is either positively, negatively or not biased to each group (n=10). **Results.** Statistical analyses suggest that interactions between bias and: 1) workload, 2) anxiety, and 3) self-control significantly affect online performance. For instance, low initial workload paired with negative bias is associated to higher peak performances (86%) than without any bias (69%). High anxiety relates negatively to performance no matter the bias (60%), while low anxiety matches best with negative bias (76%). For low CP, learning rate (LR) increases with negative bias only short term (LR=2%) as during the second session it severely drops (LR=-1%). **Conclusion.** We unveil many interactions between said human factors and bias. Additionally, we confirm them and reveal even more by using prediction models. **Significance.** This paper can serve as guideline for designing biased feedback suitable for a personality type, state, and CP in order to maximize performance and learning.

**Index Terms—**Brain-computer interface (BCI), personality traits, prediction models, feedback bias

## I. INTRODUCTION

A Brain-computer interface (BCI) can be defined as “a system that measures neural activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural neural output and thereby changes the ongoing interactions between the central nervous system and its external or internal environment” [1]. Differently from a passive BCI serving to monitor users' states through neural activity [2], active BCIs require focus to produce mental commands towards an external device, and establish a new form of control (e.g. for movement [3] or communication [4]). To achieve moderate to high performances in active BCIs the user is supposed to understand the task and to produce distinct and stable brain signals [5]. This is especially difficult for motor imagery MI-BCI, i.e., a BCI in which the machine is controlled by one's mental movements of limbs (e.g., hands, feet, tongue). To assist the user in accomplishing the task, BCI training approaches were proposed that apply various educational [6] and motivational theories [7]. The training approaches can range from using social contexts,

i.e., collaborative or competitive games [8], proprioceptive [9], immersive [10], to gamified 3D environments [11], [12], congruent with MI tasks [13], and many more [14]. In addition, some investigated whether by influencing the belief about one's current performance can in fact affect their performance in reality, e.g., by providing biased (unrealistic) feedback [15], [16], [10], [12].

While some methods such as the use of 3D game versus 2D minimalist task managed to increase both performance and user experience significantly (on average) [11], other methods produced variable results because of their dependency on human factors [17], [18], notably methods with biased feedback [15], [12]. Indeed, in [15] authors speculated that participants with initially high BCI control previously determined with calibration performance (CP) would not benefit from any type of biased feedback, while on the contrary low CP participants would. In [12] the results were inconclusive about the impact of biased feedback on performance (not significant on average), while differently from [19] where negative bias increased learning, in [10], [16] it did not. Namely, due to such variable results, authors in [10] suggested further investigation on the influence of personality types and user CP, while [16] suggested personalizing the feedback bias.

On a side note, although the mentioned methods include biased feedback (and their results were compared in [10]), their approach and interface design vary and might be another reason for such variable results. For example, perceptually continuous alterations of the classification output in real-time which resulted in mild, adaptive feedback biases were provided in [15] and [12], while in [19] and [10] authors performed absolute miss or match 30% or 90% of the time, as negative or positive feedback, respectively. Moreover, a 3D video game was presented in [12], a VR environment with proprioceptive feedback in [10], a simple 2D (ball falling in basket) task in [15], and discrete checkbox (correct/incorrect) feedback in [19]. Nevertheless, we will not investigate these issues here, such differences in interface designs and their implications in performance variations are thoroughly discussed in [20].

In order to propose training tasks or interfaces that fit each user, a variety of user traits, skills and states were identified as predictors to BCI performance. For instance, personality traits such as self-reliance positively correlates, while apprehension negatively correlates with performance [21]; and user contextual states such as motivation [22], [23], attention [24], confidence [25], sense of control [26], and state of flow (a composite optimal state of control, immersion and pleasure) [12] correlate positively to performance, while

cognitive load correlates negatively with performance but only for those with performance below 75% classification accuracy [18]. Additionally, competitiveness could be seen as a trait that predicts performance, as in sports there is evidence of a strong relationship between competitiveness, anxiety, self confidence and performance [27]; competitive players showed to give more effort when compared to non-competitive players [28], while the use of games and competitions to promote intrinsic motivation and performance has shown useful for learning various programming skills [29].

In this paper we wish to disentangle the speculations or validate results about the personality, states and CP involved in the success or failure of biased feedback. Using prediction models and standard statistical analyses, we investigate the joint effect of bias types (negative, positive and no bias) and personality traits [21], CP [15], flow state [12] and workload [18] on performance and learning. Prediction models can unveil which factors (traits, states, CP) individually predict performance and learning. Furthermore, there might be an interaction between those factors and the bias that in conjunction might predict performance or learning. Positive significant interactions would prescribe the best pairs between bias and a participant trait, state and CP that are associated with increased performance or learning. Hence, this paper is an attempt at building a guideline for matching bias to the user and maximize performance and learning.

In section II, we describe the experimental design, our bias function, signal processing and performance metrics used, as well as our prediction models. In section III we present our results including short comments; in IV we interpret more thoroughly our results, and finally in section V we provide our concluding words, and describe future works and challenges for creating an adaptive model for MI-BCI based on these results.

## II. MATERIALS AND METHODS

### A. Experimental Design

30 healthy participants were recruited (12 women, mean age: 28.56 years, SD: 6.96). The study was approved (validation number 2019-05) by the Inria ethics committee, the COERLE. We created a 3-condition-between-subject design: (1.) no\_bias (control) group in which the classifier output is not biased, (2.) positive\_bias, and (3.) negative\_bias groups in which the classifier output is positively and negatively biased in real-time, respectively. 10 participants were assigned to each group depending on the calibration scores in the first session. This would ensure the groups are balanced according to their performance. Each participant was engaged in 2 sessions (on 2 different days with a 1-to-5 days interval). For each 6-run session, participants played the Tux Racer game using left/right hand MI in order to move a virtual penguin Tux to the left/right respectively to catch fish. A run contained 40 trials (20 trials per MI class) making each course or run last 5 minutes and 25 seconds. Each 4 second long trial was visually marked by two flags on the racing course, within which a set of 8 closely arranged fish were to be caught within  $\sim 3$  seconds, see Figure 2. It was followed by a  $\sim 4$  second pause, during

which controls were deactivated, and participants could rest or adjust their position if needed. Each session lasted  $\sim 2$  hours and consisted of 3 main parts.

Part I is the experiment preparation (around 30mins), during which the participants: (a) signed the consent form; (b) filled in a demographics form; (c) watched an explanatory 3-minutes animated movie about EEG, motor imagery, and the principles of the Tux Racer game; (d) tried-out objects for motor training (see Figure 1) during which the experimenter was placing the EEG cap. Participants were not obliged to try-out any of the presented objects unless they needed a somatosensory reminder of a movement, showed to increase performance of BCI novices [30]. After the try-outs, they (e) observed their own raw EEG signals with facial movements, jaw clenching and eye blinks in real-time to increase the participant's sense of agency [31].

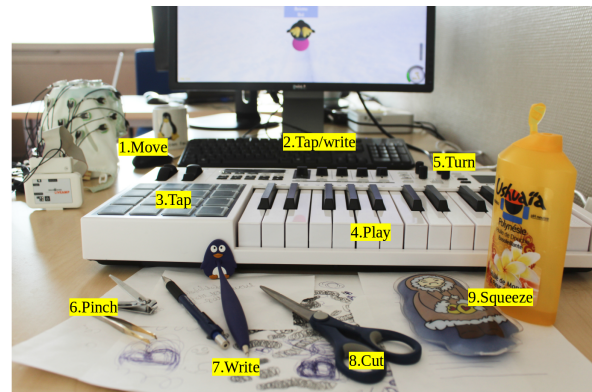


Fig. 1. Physical training on objects were suggested (non-obligatory). By demonstrating the possible movements on such objects, novice users could have a better idea of how an imagined movement can look or feel like.

Part II is the calibration phase (2 runs of 40 trials each, around 15mins), during which participants: (a) repeatedly performed their imagined movement for left or right hand at a time respectively with the left-right apparition of the fish; (b) received “fake” feedback in which Tux was controlled by a script that generated quasi-random, rather optimistic behavior (participants were told that they were not the ones controlling the game); (c) perceived Tux sliding on a pink snowboard, instead of its own belly, which was the visual indication to differentiate the calibration from the testing phase; and (d) finally, once the calibration runs were over, participants filled out a questionnaire on the subjective effort provided or workload (NASA-TLX) [32] and flow state (EduFlow) [33].

Part III is the testing phase (6 runs, around 45mins) during which the participants: (a) controlled Tux with their left-right hand motor imagery; (b) received either positively, negatively or not biased feedback; (c) filled in NASA-TLX [32] and EduFlow [33] questionnaires after each run;

Note that in each session, the machine was calibrated on training data from the same session only.

#### a) Questionnaires:

To assess the personality traits and affinities to competition, participants were asked to fill in the following questionnaires at home:

(1) Personality test 16PF5 [34], providing 16 primary scores of personality traits, and 5 global scores of personality being extroversion, anxiety, tough-mindedness, independence and self-control that are computed as linear combinations of the primary ones. Subsequent analysis were based on these 5 global scores.

(2) Revised Competitiveness Index [35], we analyze only competition enjoyment scores.

To have a better understanding of one's experience and perceived difficulty during the experiment, the participants filled 2 questionnaires after each run:

(3) EduFlow questionnaire [33] provides 4 dimensions being cognitive control, immersion, autotelism, and loss of self; the mean of all 4 dimensions we refer to as *eduflow*.

(4) NASA-TLX [32] produces one score of workload (denoted as *nasa\_score*).

## B. EEG Decoding and BCI Feedback

*a) Signal Processing:* The following equipment and software were used: (i) A 32 channel Brain Products LiveAmp (wireless amplifier), (ii) OpenViBE v2.2.0 [36] for real-time signal processing in which we used: (1) Butterworth temporal filter, whose frequency band was optimized in a subject specific-way as in [37]; (2) Common Spatial Patterns (CSP) spatial filters to reduce the 32 original channels down to 6 "virtual" channels that maximize the differences between the two motor imagery classes [38]; (3) Features computed as the EEG band power after CSP and band-pass filtering, using a 1s time window sliding every  $1/16^{th}$  seconds; (4) Probabilistic Support Vector Machine (SVM) with a linear kernel used to classify the features between left and right motor imagery (regularization parameter  $C = 1$ ). This way, the output of the SVM,  $x_i \in [0, 1]$ , indicates both the class estimate and the confidence associated with that estimation.

The CSP, SVM and frequency band selection were trained on data acquired during the calibration phase of each session.

*b) Game controls:* We used the Lab Streaming Layer (LSL<sup>1</sup>) to control a virtual joystick that in turn controls the penguin. This means that the classifier output from OpenViBE was streamed via LSL in real-time and linearly mapped onto the -45 to 45 degrees angle range the penguin would take. In between trials the controls were deactivated so that Tux could not be manipulated and participants could rest.

*c) Game design:* We modified<sup>2</sup> the open-source Tux-Racer game by creating a bobsleigh ski-course through which Tux would slide in, at constant speed. We fixed the speed so that each run would last the same amount of time. Tux would then naturally slide back to the center of the course between each trial as the controls were deactivated. We placed one row of fish close to the center (each fish providing 1 point), and a row of squids close to extremities (each squid providing 5 points). The collected points are displayed in the upper right corner, the speed and indication of position in the course are in the bottom right corner, while the time elapsed is displayed in the upper left corner, see Figure 2.



Fig. 2. Screenshot of our BCI Tux-Racer during a trial (after passing the left flag which indicates the beginning of a trial for left hand MI).

*d) Mean re-centering:* To account for the data non-stationarity, (i.e., EEG data statistics varying over time which may put the classifier out of center), we implemented the method proposed in [39]. It consists in recentering the classifier (here SVM) hyperplane over time to avoid any inclination towards one class. This way we can assure the validity of our positive/negative bias strategies. We applied this method by calculating the mean  $c$  of the classifier output for each run. As left-right classes are balanced the mean should be zero. If  $c \neq 0$  we then subtract that amount from the classifier output to center it around zero, i.e., this value becomes the new mean.

We perform re-centering after each run during testing, and calculate the mean from only the previous run (without including the calibration run). As there are many issues that could happen during a run, for instance, a faulty electrode would impact the classifier to choose one class throughout the whole run and would cause the hyperplane to undergo a drastic movement for the wrong reason, we bounded the mean estimation  $c \in [-0.5, 0.5]$ , i.e., if  $c > 0.5$  then  $c = 0.5$  or  $c < -0.5$  then  $c = -0.5$ . This method is implemented in every group, including the no\_bias group.

*e) Bias function:* For both the positive and negative\_bias groups, we used a beta cumulative distribution function (beta CDF) that maps the classifier output (which are probabilities) to a biased classifier output:

$$I_x = \frac{\int_0^x t^{a-1}(1-t)^{b-t} dt}{const}$$

where *const* is a normalization constant to ensure the total probability is 1.

In practice, we chose this function because it enables us to control the slope using its two parameters,  $a$  and  $b$ . From the identity function  $a = b = 1$  (used for the no\_bias group), we started adding and subtracting a shift value  $k$  for our bias function. We chose the parameters for positive  $a_P = 1 - k$ , and  $b_P = 1 + k$ , while  $a_N = 1 + k = b_P$  and  $b_N = 1 - k = a_P$  for the negative bias group.  $K$ -value is chosen empirically  $k = 0.33$  as a best fit for our experiment, see Figure 3. Note that beta CDF is symmetric for positive and negative biases, however it is not symmetric below and above the mean value (0.5).

## C. Evaluation Metrics

*a) Performance:* The online performance corresponds to the peak and average performance of the classifier that

<sup>1</sup><https://github.com/scenn/labstreaminglayer>

<sup>2</sup><https://github.com/jelenaLis/tux-modifs>

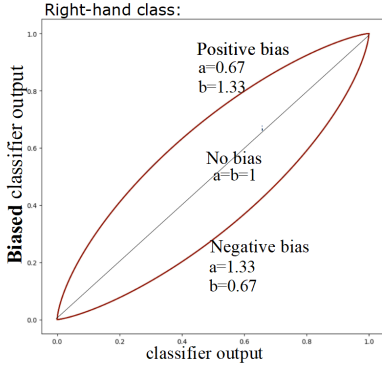


Fig. 3. Example for the right-hand class bias CD function: negative, positive and no\_bias, in which we map the classifier output ( $x$ -axis) to the biased output ( $y$ -axis), using shift value  $k=0.33$ .

controlled the video game, i.e. the highest and the mean classification accuracy over all trials' time windows, respectively. CP was computed offline with a 4-folds cross validation per run of data recorded during calibration phase.

**b) Learning:** Learning rate (LR) is the slope (coefficient) of the linear regression of online performance over runs within a session, e.g., below zero indicates a decrease in learning while above indicates an increase. LR is here session sensitive (1 value per session), as a person may not linearly continue learning from the end of the first session to the beginning of the second one. Thus, this metric is useful for observing changes between sessions, however it is less evident to predict. For this reason, in our prediction model we use a simple difference in performance between sessions, denoted as *progress*.

**c) Baseline:** During the calibration phase of each session we assessed 1 value of flow, workload and CP. As we recalibrate the system for each session, for subsequent analyses we used the assessed values (1 per session), denoted with suffix "\_baseline". However, to avoid the learning effect for progress prediction, we use only the first session values, denoted with suffix "\_reference" (1 value per participant). Thus baseline values are per session, while reference ones are per participant.

**d) High/low factors:** To have a better understanding of the interaction between bias and human factors, we separate the factors into high and low using the median value within a particular trait or state, i.e. high/low (anxiety, extroversion, independence, tough-mindedness, self-control), and high/low (workload, flow, CP).

#### D. Prediction models

Let online performance be  $Y^n$  for  $n$  subjects, using  $X^{n \times p}$  covariates for  $p$  predictive factors. We are interested in the interactions between three bias types  $r_k \in \mathbb{B}$ ,  $k = 0, 1, 2$  on one hand, and traits (5 global personality traits, competition enjoyment), baseline states (CP, flow and workload "\_baselines"),  $s_j \in \mathbb{R}$ ,  $j = 1, \dots, 9$  on the other. Finally, the interaction between  $r_k$  and  $s_j$  or product  $r_k s_j = x_{9k+j}$  yields  $p = 27$  predictors  $X^{n \times p}$ . We implement an Elastic-Net regression [40], which combines both Ridge (Tikhonov regularization) and Lasso

(Least Absolute Shrinkage and Selection Operator) regression, as follows.

$$\beta_{net} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} (\|Y - X\beta\|_2^2 + \lambda [\frac{1}{2}(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1])$$

Where  $\beta$  represent regression weights,  $\lambda$  is the regularization or shrinkage parameter, while the parameter  $\alpha$  balances the respective contribution of the two penalty terms  $\|\beta\|_2^2$  and  $\|\beta\|_1$ . Strictly speaking,  $\|\cdot\|_2^2$  indicates the  $l_2$  norm while  $\|\cdot\|_1$  indicates the  $l_1$  norm. If  $\alpha = 0$  then it boils down to Ridge regression with  $l_2$  penalty, otherwise if  $\alpha = 1$  it amounts to Lasso regression with  $l_1$  penalty.

Two or more predictors are said to interact if their combined effect is different (lesser or greater) than the sum of the effect of each predictor taken separately. Then our model for predicting performance  $y_i$  for each subject  $i$ :

$$y_i = \beta_0 + \sum_{k=0}^2 r_{i,k} \sum_{j=1}^9 \beta_{9k+j} s_{i,j}$$

where  $\beta_{9k+j}$  are the coefficients (from 1 to 27) of the interactions  $r_{i,k}$  and  $s_{i,j}$ . Knowing that  $r_{i,k} \in \mathbb{B}$  is a Boolean, categorical predictor, there exists only one  $k$  among 0, 1 and 2 such that  $r_{i,k} = 1$ . In other words, there can be only one bias type per participant  $i$ , the other two are equal to zero. Note that the same procedure is performed for the prediction of progress, except that  $s_{i,j}$  baseline states or "\_baselines" are replaced with "\_references", e.g.  $s_{i,1}$  is CP reference.

To optimize the prediction model, we perform a Leave One Subject Out (LOSO) cross-validation using the R package `cva.glmnet`, from `glmnetUtils`<sup>3</sup>, which uses a coordinate descent algorithm that, for each  $\alpha \in [0, 1]$  (10 values by default), generates an entire path of solutions in  $\lambda$  (100 values). We performed a nested cross-validation LOSO in order to have different sets for training and testing during parameter search, and to avoid over-fitting. This means that we performed a  $n$ -fold cross validation in the outer loop for model evaluation and a  $(n-1)$ -fold cross validation in the inner one for hyperparameter  $(\alpha, \lambda)$  selection.

In order to assess the prediction power of our models we compared their performance against thousand random models, where we kept hyperparameter values and user's factors but shuffled the outcome variable (either performance or progress). We could thus compare statistically the observed error rate with error rates derived from our null hypothesis. In order to get the final coefficients of the regression, we run the model using the selected parameters  $\alpha$  and  $\lambda$  on the whole dataset. Finally, to assess the significance of features within the model, it is compared with the random model with a thousand permutations using R package `eNetXplorer` [41]. This statistical test based on permutation is robust to repeated measurements, in our case the several performance scores per subject.

For practical reasons, we predict online performance per participant using data from each run (12 values per participant). We indeed used performance over runs so that

<sup>3</sup><https://cran.r-project.org/web/packages/glmnetUtils/vignettes/intro.html>

eNetXplorer can compute enough permutations to test the significance with  $p < 0.01$ . Nevertheless, almost identical coefficients and RMSE were found for the model of performance averaged per session. Results' similarities confirm the insensitivity of permutations to dependent (within subject) factors. However, for the sake of readability, we do not report those model's results.

### E. Statistical Analysis

Our statistical analysis involves ANOVA tests for determining the differences between our 3 bias groups. All the corresponding p-values were corrected for multiple comparisons with false discovery rate (FDR) [42], with the threshold being set to  $p < 0.05$ . To avoid repetition, all results are presented after the p-values were corrected with FDR.

As for evaluating the validity of prediction models, we use the standard metric such as root mean squared error (RMSE), and the coefficient of determination  $R^2$  (adjusted for multiple predictors).

## III. RESULTS

### A. Group balance verification

When performing 1-way ANOVA (independent variable: group, dependent: either calibration reference, dimensions of EduFlow score from calibration, NASA-TLX score from calibration, or personality traits) there is no significant difference between groups. This could indicate that the groups were balanced considering their baseline performance, flow states and workload at the beginning of the experiment.

### B. Differences between groups within sessions

In the following, we use workload, eduflow and CP from both calibration phases (nasa\_baseline, flow\_baseline, calibration baseline).

*a) Performance:* When performing 3-way ANOVAs (independent variables: group, session and high/low factors, dependent variable: either peak or average performance), there is a significant effect ( $p < 0.001$ ) of high/low calibration performance on both peak and average online performance, meaning that e.g. low calibration performance would yield low online performance, no matter the bias.

We found a significant interaction: bias $\times$ high/low workload, for both peak ( $p = 0.036$ ), and average performance ( $p = 0.022$ ), see Fig. 4. This suggests that participants with initially low workload (low nasa\_baseline) benefit most from a negative bias (peak $\sim$  85%) and least from no\_bias (peak $\sim$  67%). On the other hand, high workload participants do not require any bias to gain reasonably high performances (peak $\sim$  78%).

There was a significant interaction: bias $\times$ high/low anxiety for average ( $p = 0.031$ ), and a tendency for peak performance ( $p = 0.080$ ), see Fig. 5. This would suggest that low anxiety participants would benefit most from a negative bias (avg $\sim$  77%) and least from no\_bias (avg $\sim$  61%). On the other hand, highly anxious participants have low performance no matter the bias (avg $\sim$  59%).

Finally, there was a significant interaction: bias $\times$ high/low self-control for average ( $p = 0.037$ ), and for peak performance ( $p = 0.042$ ), see Fig. 6. This result suggests that high self-controlled participants have higher performances with negative feedback (peak $\sim$ 82%), than with no\_bias (peak $\sim$ 65%) which is the inverse for low self-controlled participants.

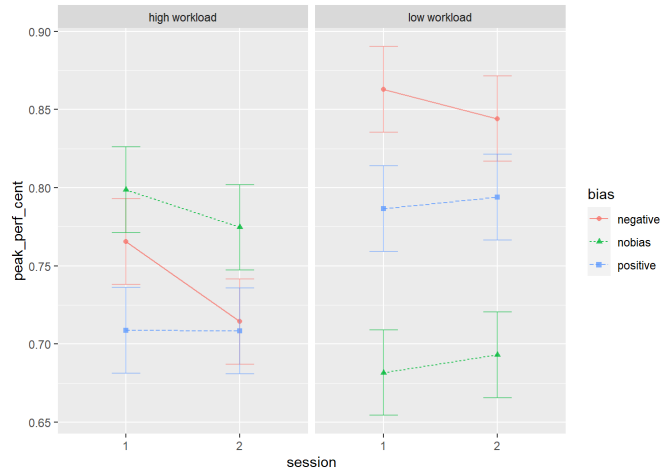


Fig. 4. Significant interaction between high/low workload and bias ( $p < 0.05$ ) for peak performance within 2 sessions.

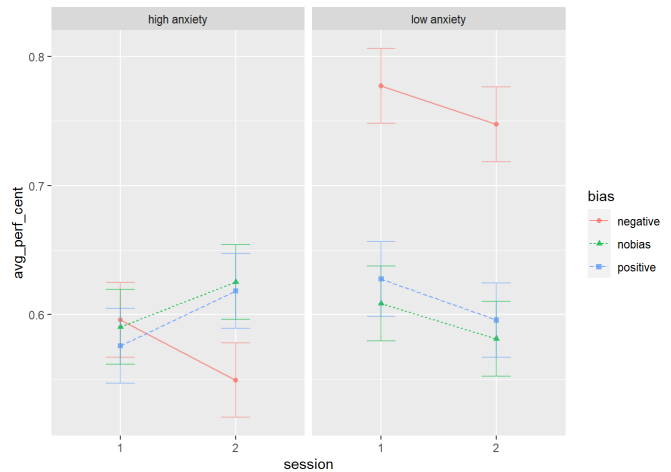


Fig. 5. Significant interaction between high/low anxiety and bias ( $p < 0.05$ ) for average performance within 2 sessions.

*b) Learning rate:* With a 3-way ANOVA (independent variable: group (bias), session, high/low factor; dependent variable: learning rate) we get significant interactions: bias $\times$ session ( $p=0.05$ ), high/low CP $\times$ session ( $p < 0.05$ ); and high/low CP $\times$ bias $\times$ session ( $p < 0.01$ ), see Figure 7. This result indicates that CP, and bias can directly influence learning within sessions.

Negative bias increases learning rate within first session for those with low CP, but already in the second session it severely impedes learning rate, from LR $\sim$ 2% to LR $\sim$ -1%. Positive bias seems to have a generally negative influence on learning (mostly below zero). While no\_bias seems to be the best fit for high CP, although for only one session (from LR $\sim$ 1.5% to LR $\sim$ 0.4%).

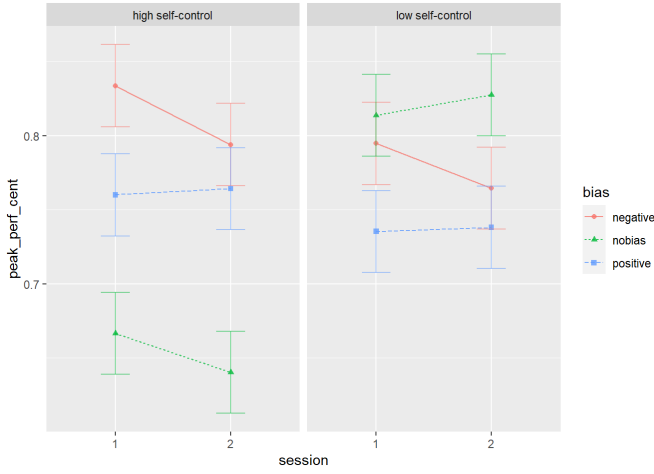


Fig. 6. Significant interaction between high/low self-control and bias ( $p < 0.05$ ) for peak performance within 2 sessions.

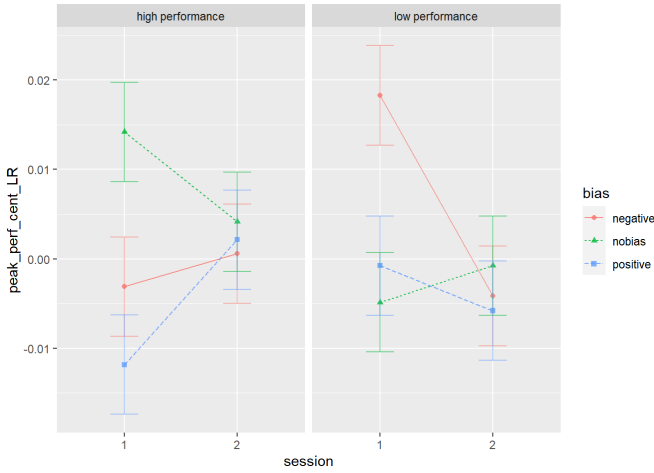


Fig. 7. Significant interaction between high/low calibration performance and session ( $p = 0.027$ ) and bias all together ( $p = 0.002$ ) for learning rate of peak performance.

### C. Prediction Models

Note that performance is presented in percentages, and that predictors (traits, flow and workload and CP) were normalized with min-max prior the elastic-net to enable comparison.

1) *Prediction of Online Performance*: Our model investigates only interactions between bias and remaining predictors that jointly affect performance. Parameters selected by elastic-net are  $\alpha = 0.50$ , with shrinkage  $\lambda = 0.060$  giving an error  $RMSE = 10.10$  and deviation ratio  $R^2 = 0.54$  for average performance, and  $\alpha = 0.69$  with  $\lambda = 0.050$  and  $RMSE = 10.72$  and  $R^2 = 0.53$  for peak performance. When compared to the random model they are significantly better than chance  $p < 0.0001$  for both average and peak.

Eighteen interactions were selected by our model for both peak and average performance, see Figure 8. Novel interactions are revealed in our prediction model with: (1) extroverted participants who might benefit from negative bias to attain better performance, ( $p < 0.05$ ); (2) competitive ones might benefit from no\_bias ( $p < 0.01$ ), while their performance can be

impeded with negative bias ( $p < 0.05$ ), (3) tough-minded ones seem to worsen their performance without any bias ( $p < 0.01$ ), (4) participants initially in flow (baseline state) do not seem to require a biased feedback ( $p < 0.01$ ).

Interaction model	Average perf Intercept (47.13)			Peak perf Intercept (58.22)			
	Bias	Neg bias	Pos bias	No bias	Neg bias	Pos bias	No bias
<b>Predictors</b>							
Calib baseline		32.6**	24.2**	4.2	35.0**	35.0**	4.1
Anxiety		-14.4**	-9.8**	1.1	-11.4**	-16.5**	1.8
Self-control			10.2**	-14.0**	0.05	10.1*	-11.1**
Extroversion		10.6*	1.2		6.8*		
Competition enjoyment		-5.1		14.0**	-5.0*		18.1**
Independence			-6.3	2.5		-6.5	2.8
Tough mindedness		0.9		-7.6*	1.4		-15.2**
Flow baseline				8.4*			10.9**
Nasa baseline				14.0**			16.8**

Fig. 8. Coefficients of predictors that significantly interact with bias (negative, positive and no bias). There are two intercepts, one for each performance. Significant coefficients (\* for  $p < 0.05$ , \*\* for  $p < 0.01$ ).

2) *Prediction of Progress*: Progress calculated from peak performance can be predicted better than chance ( $p < 0.0001$ ), which is not the case for average performance. Model parameters  $\alpha = 0.18$  and  $\lambda = 0.001$  provide  $RMSE = 3.05$  and  $R^2 = 0.87$ . Twenty-seven interactions were selected, however we could not test the significance of their coefficients with eNetXplorer due to the lack of data (1 progress value) per participant.

Interaction model	Progress (peak) Intercept (-34.64)			
	Bias	Negative	Positive	No bias
<b>Predictors</b>		Coefficients		
Calibration reference		8.9	16.7	7.8
Anxiety		13.7	-15.3	-0.2
Self-control		-46.2	16.2	25.3
Extroversion		37.3	25.4	-12.5
Competition enjoyment		45.6	-12.2	35.9
Independence		-16.6	45.1	44.6
Tough mindedness		2.1	-9.5	-30.4
Flow reference		12.3	-3.9	34.5
Nasa reference		-18.1	16.5	-32.6

Fig. 9. Predicting progress through the interaction between bias types on one hand, and traits, reference states on the other, and their corresponding coefficients.

## IV. DISCUSSION

### A. Results from ANOVAs

It seems that in most cases, negative bias can be beneficial for performance and learning, but for one session only, as it impedes them already in the second session, and significantly so for learning of low CP participants (see Fig. 7). This result implies the potential of adaptive bias within sessions for performance and learning maximization.

As shown previously in [43], biased feedback directly influences the participant's learning and flow state (cognitive control), however it does not influence performance due to high population variability in calibration performance, personalities and initial, baseline states (e.g. flow and workload) among others. When the participants are divided by high/low factors, we observe the following results.

**Workload.** Those who have high workload baseline can benefit from no\_bias feedback (peak\_perf  $\sim 77\%$ ), suggesting that these participants do not need such feedback assistance as they are engaged in the task from the beginning. On the other hand, those with low workload baseline can outperform those with initially high workload when presented with altered feedback, specifically negative one (peak\_perf  $\sim 85\%$ ), from Fig. 4. This suggests either that participants who start with low workload levels might attain high CP or that they can be motivated to engage more strongly when faced with a seemingly more difficult task. Further investigation is necessary to confirm.

**Anxiety.** Those with low anxiety can significantly increase performance with negatively biased feedback (avg\_perf  $\sim 85\%$ ), suggesting that a more difficult task can seem more motivating for them; while those with high anxiety perform poorly given any feedback bias (avg\_perf  $< 65\%$ ), from Fig. 5.

**Self-control.** High self-controlling participants seem to benefit from negative bias (peak\_perf  $\sim 85\%$ ), while low self-controlling ones do not require any bias (peak\_perf  $\sim 85\%$ ). Self-controlling participants are highly rule conscious, perfectionists, while lacking liveliness and abstractedness. This result corroborates with [21] where abstractedness is positively correlated with performance. Indeed, those who are not self-controlling or manifest abstractedness in their personality do not require an altered feedback to attain high performances, from Fig. 6. On the other hand, a more difficult task might enable them to switch their attention from the rules and perfectionism, and focus on the task at hand.

**Learning rate.** Results suggest that those with low CP can compete with high CP (LR=1.5%) when presented with negative feedback (LR=2%), but only for one session (short term), see Fig. 7.

## B. Results from prediction models

Our prediction models confirm results from ANOVAs but unveil even more interactions between predictors and bias, see Fig. 8.

**Confirming ANOVAs.** Calibration baseline is confirmed as a strong predictor of performance as hypothesized in [15], disregarding the feedback bias. On the contrary, anxiety is detrimental for performance no matter the bias valence, see ANOVA Fig. 5. No\_bias seems to degrade performance the most for self-controlling participants ( $p < 0.01$ ), see ANOVA Fig. 6. Our prediction model also confirms that high workload participants do not require a bias (no\_bias) to reach high performance ( $p < 0.01$ ), see ANOVA Fig. 4, and our model Fig. 8.

**New interactions.** Extroverted participants might benefit from negative bias to attain better performance, ( $p < 0.05$ ). Being in positive correlation with CP and online performance (cf.

Appendix V-A3) suggests that extroverted participants attain high CP, and high performances. In respect with the Flow theory, in which an easy task might bring boredom and a difficult one frustration, we can concur that a more difficult task can bring higher motivation to these participants, and thus higher performance. On the other hand, given the negative correlation between CP and workload, it implies that less effort or workload is required during testing for high CP participants, and given the ANOVA of workload Fig. 4 shows that negative bias is useful for low workload participants; thus it might be useful for extroverted ones as well. Clearly, further investigation is needed to confirm this claim.

Competitive participants do not seem to need altered feedback to attain high performance, ( $p < 0.01$ ) as they are already motivated from start, while negative bias can impede their performance ( $p < 0.05$ ). Competitive players seem to give more effort [28], or have higher workload. Interestingly, those with high workload also benefit from no\_bias and not from a negative one (ANOVA Fig. 4).

Participants initially in flow (baseline) do not seem to require a biased feedback to attain high performance ( $p < 0.01$ ), this can confirm that neither too easy (positive bias) or too difficult (negative bias), but the middle (no\_bias) tasks are useful for performance when in flow; confirming the positive correlation between flow and performance [12], cf. Appendix V-A1.

	Positive bias	Negative bias	No bias
Performance	CP* <sup>A</sup> Self-Control* <sup>A</sup> Extroversion	CP* <sup>A</sup> Extroversion* Tough mindedness Self-Control <sup>A</sup>	Workload* <sup>A</sup> Competition enjoyment* Flow* Anxiety Independence CP <sup>A</sup>
	Anxiety* <sup>A</sup> Independence	Anxiety* <sup>A</sup> Competition enjoyment*	Self-control* <sup>A</sup> Tough mindedness*
Progress	Independence Extroversion CP Workload Self control	Competition enjoyment Extroversion CP Anxiety Flow Tough mindedness	Independence Competition enjoyment Flow Self control CP <sup>A</sup>
	Anxiety Competition enjoyment Tough mindedness Flow	Self control Independence Workload	Workload Tough mindedness Extroversion

Fig. 10. Preliminary guidelines for fitting biased feedback to each BCI user. Positive interaction is in green, while negative in red. (A) confirms ANOVA results, while, for performance significant predictions contain (\*).

Interestingly anxiety together with negative bias can be useful for progress, but detrimental for performance. Indeed, as flow can be useful for performance and not learning (negative correlation between flow and learning rate, cf. Appendix V-A2), it is possible that anxiety (in a way opposite from flow) can be useful for learning.

We summarize the results from all our prediction models to ease their understanding in Figure 10. However, results that were not shown significant in the prediction models are not to be accepted as absolute truth, especially with RMSEs as large. Prediction models can merely give an intuition on potential biases to be avoided or favoured for certain personalities, states and CP. Clearly, further investigation is necessary to validate



some newly found predictors. From our results we derive a preliminary guideline for providing optimal bias for each user that would maximize performance and learning.

## V. CONCLUSION

The literature suggests that benefits of biased feedback vary across users, or that there is an interaction between bias types and human factors. In this paper, we bring clarity about the dependencies of personalities, states and CP on bias types (negative, positive and no\_bias), and their joint effect on performance and learning. In other words, we unveil the combinations of a bias type and said human factors that might increase performance and learning. This paper can be used as a guideline when designing biased feedback so that one can be able to fit a bias type for each user. Furthermore, an adaptive bias between sessions has emerged as a possible beneficial solution to maximize performance and learning. In the future, these predictors for optimal bias can provide prior information for an adaptive framework that could adapt the bias automatically for each user.

## ACKNOWLEDGMENT

This work received support from Inria project lab BCI-LIFT and from the European Research Council with project BrainConquest (grant ERC\_2016\_STG\_714567).

## APPENDIX

### A. Correlation results

We found many pertinent correlations between traits, states and performance, LR and CP, as follows, see simplified Figure 11.

Correlations		
Performance	Flow: Control, Selflessness, Extroversion	LR
	Anxiety	CP
	Workload	Extroversion
	Flow, Control, Selflessness	Workload, Flow, Anxiety

Fig. 11. Correlations with performance, learning rate (LR), and calibration performance (CP) representing only valence – positive (green) or negative (red).

1) *Performance*: We correlate flow and workload scores per run with online performance per run. There are significant correlations between average, peak performances respectively and: a. Eduflow score ( $p < 0.0001$ ,  $r=0.22$ ,  $r=0.23$ ), with its dimensions b. Control ( $p < 0.0001$ ,  $r=0.26$ ,  $r=0.25$ ), c. Selflessness ( $p = 0.012$ ,  $r=0.15$ ,  $r=0.14$ ); and traits e. Extroversion ( $p < 0.05$ ,  $r=0.40$ ,  $r=0.36$ ); f. Anxiety ( $p < 0.01$ ,  $r= -0.42$ , and  $r= -0.39$ ). Finally, autotelism correlates only with peak performance ( $p = 0.005$ ,  $r=0.16$ ), while the 2<sup>nd</sup> dimension of flow, immersion, and other traits do not correlate with any online performance. Namely, workload does not correlate significantly with performance for all participants on average. Based on the literature [18], we performed an additional correlation to verify whether workload of those with low online performance (peak and average < 75%) is indeed negatively correlated with online performance. However, in our case it remains not significant.

2) *Learning Rate*: Learning rate and EduFlow dimensions (control and loss of self) correlate negatively (both with  $p < 0.01$ ), especially the sense of control ( $r = -0.20$ ). In contrast, workload correlates positively with learning ( $r= 0.13$ ,  $p < 0.05$ ). We found no significant correlations between traits and learning rate.

3) *Calibration Performance*: Interestingly, there are significant correlations between calibration reference and: a. Anxiety ( $p < 0.0001$ ,  $r=-0.37$ ), b. Extroversion ( $p < 0.0001$ ,  $r=0.33$ ), and c. Eduflow reference ( $p < 0.0001$ ,  $r=-0.26$ ). Finally, nasa score correlates negatively with calibration baseline ( $p=0.03$ ,  $r=-0.11$ ).

**Interpretation:** As expected, flow with its dimensions correlate positively with performance [12], while anxiety (containing tension) correlates negatively with online performance [21], with CP as well. This shows how anxiety is detrimental for performance at any moment during the experiment.

Flow is negatively correlated with learning rate and with CP. This is indeed quite interesting but contrary to positive impact of flow in educational games [44]. Flow state represents a state of pleasure, immersion in the present moment, thus BCI performances in short term can benefit from such a state, but not necessarily learning. As for CP, it is possible the participants were "too relaxed" and did not engage enough with the task. This might be what differentiates BCI tasks from regular ones, the need for a higher attentional focus.

As any experiment is a social event having to interact with experimenters in a new environment, it seems quite plausible that those participants who are extrovert have higher CP, and higher online performance.

Surprisingly, workload does not correlate significantly with online performance for all participants on average, even when divided into workload of high/low performance participants. This fails to confirm the results from [18] where workload negatively correlated with performance for participants with classification accuracy below 75%. However, the difference can be due to numerous reasons, e.g. the performance metrics differ from our online performance. Moreover, we found that workload score (acquired after each run) correlates negatively with calibration baseline (CP from each session). This makes sense, as those with low CP will give more effort during testing phase to compensate for the wrongly calibrated machine, while those who managed to perform well during calibration, they do not need to put as much effort later-on. Clearly, further investigation is needed to better understand the evolution of workload depending on the bias and CP.

## REFERENCES

- [1] J. R. Wolpaw, "Brain-computer interfaces," in *Handbook of Clinical Neurology*. Elsevier, 2013, vol. 110, pp. 67–74.
- [2] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general," *Journal of neural engineering*, vol. 8, no. 2, p. 025005, 2011.
- [3] J. d. R. Milan and J. M. Carmena, "Invasive or noninvasive: Understanding brain-machine interface technology [conversations in bme]," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. 1, pp. 16–22, 2010.

- [4] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.
- [5] F. Lotte and C. Jeunet, "Defining and quantifying users' mental imagery-based bci skills: a first step," *Journal of neural engineering*, vol. 15, no. 4, p. 046030, 2018.
- [6] J. Keller, "Challenges in Learner Motivation: A Holistic, Integrative Model for Research and Design on Learner Motivation," in *New Educational Paradigm for Learning and Instruction*, 2010, pp. 1–18.
- [7] J. Nakamura and M. Csikszentmihalyi, "The Concept of Flow," in *Handbook of positive psychology*, C. R. Snyder and S. J. Lopez, Eds. New York: Oxford University Press, 2002, pp. 89–105.
- [8] L. Bonnet *et al.*, "Two brains, one game: Design and evaluation of a multiuser bci video game based on motor imagery," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 5, no. 2, pp. 185–198, 2013.
- [9] N. Braun *et al.*, "Embodied neurofeedback with an anthropomorphic robotic hand," *Scientific reports*, vol. 6, p. 37696, 2016.
- [10] M. Alimardani *et al.*, "Effect of biased feedback on motor imagery learning in BCI-teleoperation system," *Frontiers in systems neuroscience*, vol. 8, no. April, p. 52, 2014.
- [11] R. Ron-angevin, "Brain-computer interface: Changes in performance using virtual reality techniques," vol. 449, pp. 123–127, 2009.
- [12] J. Mladenović *et al.*, "The impact of flow in an EEG-based brain computer interface," *International Graz BCI conference*, 2017.
- [13] E. Christophe *et al.*, "Evaluation of a congruent auditory feedback for motor imagery BCI," *International BCI Meeting*, 2018.
- [14] A. Roc *et al.*, "A review of user training methods in brain computer interfaces based on mental tasks," *Journal of Neural Engineering*, 2020.
- [15] A. Barbero and M. Grosse-Wentrup, "Biased feedback in brain-computer interfaces," *Journal of neuroengineering and rehabilitation*, vol. 7, p. 34, 2010.
- [16] I. N. Angulo-Sherman and D. Gutiérrez, "Effect of different feedback modalities in the performance of brain-computer interfaces," in *2014 International Conference on Electronics, Communications and Computers (CONIELECOMP)*. IEEE, 2014, pp. 14–21.
- [17] D. J. McFarland *et al.*, "Eeg-based communication and control: short-term role of feedback," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 1, pp. 7–11, 1998.
- [18] Z. Emami and T. Chau, "The effects of visual distractors on cognitive load in a motor imagery brain-computer interface," *Behavioural brain research*, vol. 378, p. 112240, 2020.
- [19] M. Gonzalez-Franco *et al.*, "Motor imagery based brain-computer interface: A study of the effect of positive and negative feedback," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 6323–6326.
- [20] J. Mladenovic, "Standardization of protocol design for user training in EEG-based brain-computer interface," *Journal of Neural Engineering*, 2020.
- [21] C. Jeunet *et al.*, "Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns," *PLoS ONE*, 2015.
- [22] S. Kleih *et al.*, "Motivation modulates the P300 amplitude during brain-computer interface use," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1023–1031, jul 2010.
- [23] E. M. Hammer *et al.*, "Psychological predictors of smr-bci performance," *Biological psychology*, vol. 89, no. 1, pp. 80–86, 2012.
- [24] M. Grosse-Wentrup, "Fronto-parietal gamma-oscillations are a cause of performance variation in brain-computer interfacing," in *2011 5th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2011, pp. 384–387.
- [25] F. Nijboer *et al.*, "An auditory brain-computer interface (bci)," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 43–50, 2008.
- [26] M. Witte *et al.*, "Control beliefs can predict the ability to up-regulate sensorimotor rhythm during neurofeedback training," *Frontiers in human neuroscience*, vol. 7, no. August, p. 8, 2013.
- [27] J. J. Martin and D. L. Gill, "The relationships among competitive orientation, sport-confidence, self-efficacy, anxiety, and performance," *Journal of Sport and Exercise Psychology*, vol. 13, no. 2, pp. 149–159, 1991.
- [28] A. L. Snyder *et al.*, "Virtual and live social facilitation while exergaming: competitiveness moderates exercise intensity," *Journal of Sport and Exercise Psychology*, vol. 34, no. 2, pp. 252–259, 2012.
- [29] J. C. Burguillo, "Using game theory and competition-based learning to stimulate student motivation and performance," *Computers & education*, vol. 55, no. 2, pp. 566–575, 2010.
- [30] A. Battison *et al.*, "Effectiveness of subject specific instruction on mu-based brain-computer interface training," in *2015 41st Annual Northeast Biomedical Engineering Conference (NEBEC)*. IEEE, 2015, pp. 1–2.
- [31] R. Vlek *et al.*, "Bci and a user's judgment of agency," in *Brain-Computer-Interfaces in their ethical, social and cultural contexts*. Springer, 2014, pp. 193–202.
- [32] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.
- [33] J. Heutte *et al.*, "The EduFlow Model: A Contribution Toward the Study of Optimal Learning Environments," in *Flow Experience*. Cham: Springer International Publishing, 2016, pp. 127–143.
- [34] R. B. Cattell and H. E. P. Cattell, "Personality structure and the new fifth edition of the 16pf," *Educational and Psychological Measurement*, vol. 55, no. 6, pp. 926–937, 1995.
- [35] J. Houston *et al.*, "Revising the competitiveness index using factor analysis," *Psychological Reports*, vol. 90, no. 1, pp. 31–34, 2002.
- [36] Y. Renard *et al.*, "OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain-Computer Interfaces in Real and Virtual Environments," *Presence*, vol. 19, no. 1, pp. 35–53, 2010.
- [37] B. Blankertz *et al.*, "Optimizing spatial filters for robust eeg single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2007.
- [38] H. Ramoser *et al.*, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans Rehabil Eng*, vol. 8, no. 4, pp. 441–446, 2000.
- [39] C. Vidaurre *et al.*, "Toward unsupervised adaptation of lda for brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 587–597, 2010.
- [40] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [41] J. Candia and J. S. Tsang, "enetexplorer: an r package for the quantitative exploration of elastic net families for generalized linear models," *BMC bioinformatics*, vol. 20, no. 1, p. 189, 2019.
- [42] W. S. Noble, "How does multiple testing correction work?" *Nature biotechnology*, vol. 27, no. 12, pp. 1135–1137, 2009.
- [43] J. Mladenović *et al.*, "Biased feedback influences learning of motor imagery BCI," in *International BCI Meeting*, 2020.
- [44] W. Admiraal *et al.*, "The concept of flow in collaborative game-based learning," *Computers in Human Behavior*, vol. 27, no. 3, pp. 1185–1194, 2011.