

# Reply to Reviewer's Comments on "Training a gaming agent on brainwaves"

We are grateful to the reviewer for pointing out relevant issues in our manuscript.  
In the following, we discuss how we dealt with each raised issue.

.....

## REVIEWER #1 TRANSCRIPT:

This paper proposes the use of a gaming agent to be trained using Reinforcement Learning (RL) as well as the feedback obtained from EEG brainwaves of human critic observers. The idea of using brain patterns for boosting ML is highly interesting and well suited for the journal. Although it is not a new concept, it is nicely re-purposed for online use-case.

However, I am not sure if there is something new in the paper compared to the 2010 proof-of-concept (Iturrate et al.: Robot Reinforcement Learning using EEG-based reward signals). The presentation style seems a bit confusing and needs serious improvements. Background is very limited (some are incorporated in the introduction) and substantial work is required. There has been a lot of work in BCIs over the past decade and some in the area of games.

From the scientific part, the results are not so convincing. In particular, the proposed idea does not seem to be too practical method. The classification does not provide something significant as it stands at the moment. The sample used is also quite small and the complexity of the game is limited.

Nevertheless, the paper could be improved in many ways. An obvious point would be to improve the classifiers. Apart from this, I would suggest that authors would look at ways of manipulating the stimulus, stimulus presentation or giving out incentives to participants.

## REVIEWER #2 TRANSCRIPT:

The paper is interesting and relevant. However in my opinion it needs revision for purposes of clarification and to improve the contribution. Please see attached document for my comments.

Training a Gaming Agent on Brainwaves There are too many grammatical errors to list individually. //Does the header refer to a general template? JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015 //The abstract read more like an introduction, rather than an indication of the work undertaken for the paper. 25 "Results show that there is an effective transfer of information and that the agent learns successfully to solve the game efficiently." //This is too vague. What are the outcomes of the study? 39 This information is used to make a gaming agent improves its operational performance using electroencephalography (EEG) signals as feedback of the performed task, obtained from an observational human critic. //There is a grammar issue. I don't understand this sentence. 54 RL should be defined in main body of the paper. Also the term 'agent' needs to be defined/clarified. How does it relate to the Game Manager from Figure 1? 17 col 2 Recently, this technique has seen a come-back. //this is not scientific language 58 col 3 The precision of 25.125 is inappropriate P2, 46 col 2 What is "state information"? P3, 2 red normally indicates an error P3, 11 I assume that the "observational human critic" is the player/participant/subject/human observer. This role should be clarified and terminology used consistently. P3, 41 define MNE P3, 56 Thus, each epoch is composed of a matrix 500 x 8. // add channels P4, 32 Hence, following the iterative procedure based on Equation 1, the Q-Table is updated in each iteration. After the algorithm finishes iterating through all the training episodes, the Q-Table is stored to test the performance of the agent. // Will the game always terminate? How long does the game take? Does smooth progression toward finish affect the err potential? P4 Fig 3. Is the chance score = 0.5?

P4 the labels referred to in Fig 5 should correlate with the test here, i.e., A= subject 1 etc. P4 referring to Fig 6; a comparison ROC curve with subject 1 would be more interesting

P5, 34 what is meant by “experiences”? P5, 29 col 2 – remove Average steps per Q-Table legend. P5 50, col 2 The collected data show that ErrP signals can in fact be classified and used to train an agent effectively. // how has effectiveness been determined here? Page 6 Fig 10 the text is not legible, it should be improved. Page 6, 40 “However, even though this implies that the agent misses frequently that an action taken is wrong, this is not hindering the overall performance and the agent is still learning.” // Your results show that this is subject dependent P6, 56 Results show that training a classifier with data of one subject, but using it to classify the events of experiences of another subject does not lead to an improvement on the performance of the agent. // could a pre-trained generic classifier provide a better initial state, subsequently trained with observer data to converge more quickly? Are there differential error potential for up, down, left, right?

#### REVIEWER #3 TRANSCRIPT:

Comments to the Author The following work shows the usage of ErrPs for training a gaming agent using reinforcement learning. Authors attempt different conventional classification approaches, as well as intersubject classification.

It is unclear to me the benefit of evaluating single subject to single subject offline classification accuracy. Please elaborate on this. Since different subjects attained different performance, all combinations of subjects used for training and testing should be inspected for the 1:1 evaluation setup. Maybe classification accuracy could be reported at different steps (depending on the amount of training data – gradual increasing the number of subjects) What is the ratio of hit/no hit segments of action (epochs)? Page 3, column 2, lines 4-9. It is unclear at this stage what action from the starting point would generate an ErrP. moves-¿ move Page 3, column 2, lines 11-13. Has the MinMax Scaler been applied in any other BCI related study or elsewhere? I think including a reference would be useful. Page 4. Figure 3. I would suggest adding another set of bars for the grand average classification scores. Page 4, Figure 4. This figure is redundant and could be removed. Page 5, Figure 5. The titles of the subplots would be more descriptive when replacing the letters with the corresponding subjects number. Page 5, column 1, line 52, Not -¿no Page 5, column1, line 55, accumulative-¿cumulative Page 6, Figure 8. It is unclear to me what is the benefit of showing both A and B. Since the behavior is similar in both cases. Page 6, Figure 10. This figure is, in my opinion, incomplete. What is the reason for showing this subset of subjects? I suggest showing the confusion matrices of all subjects and/or their average. Page 6, column 2, line 40. show -¿ shows

#### Reviewer 1 General Comments

This paper proposes the use of a gaming agent to be trained using Reinforcement Learning (RL) as well as the feedback obtained from EEG brainwaves of human critic observers. The idea of using brain patterns for boosting ML is highly interesting and well suited for the journal. Although it is not a new concept, it is nicely re-purposed for online use-case.

However, I am not sure if there is something new in the paper compared to the 2010 proof-of-concept (Iturrate et al.: Robot Reinforcement Learning using EEG-based reward signals). The presentation style seems a bit confusing and needs serious improvements. Background is very limited (some are incorporated in the introduction) and substantial work is required. There has been a lot of work in BCIs over the past decade and some in the area of games.

We conducted a literature review and updated the introduction accordingly. We highlighted what we think is our contribution, which is a simple game that can be used to trigger the ErrP response in a very simple scenario that can be used to train. Additionally, regarding the results that we got, we think it is important to emphasize the cumulative contribution of different subjects enhances the performance of an agent even though the classification is not good.

From the scientific part, the results are not so convincing. In particular, the proposed idea does not seem to be too practical method. The classification does not provide something significant as it stands at the moment. The sample used is also quite small and the complexity of the game is limited.

Issues raised by the Reviewer are accurate and timely described. We aim to find a suitable game that at the same time can be implemented by RL and that triggered the ErrP response. AFAIK we do didnt find any experiment that tackles the generation of ErrP for games in this way. Most of the work replicates the structure of Iturrate experiment in robotics but none on a simple game plataform as the one described here.

Nevertheless, the paper could be improved in many ways. An obvious point would be to improve the classifiers. Apart from this, I would suggest that authors would look at ways of manipulating the stimulus, stimulus presentation or giving out incentives to participants.

Due to COVID crisis we are unable to perform further experimentation.

#### *Reviewer 2 General Comments*

The paper is interesting and relevant. However in my opinion it needs revision for purposes of clarification and to improve the contribution. Please see attached document for my comments.

Training a Gaming Agent on Brainwaves There are too many grammatical errors to list individually.

We truly appreciate this level of detail in all the comments and the requests for information.

The value provided by the comments in huge. Thanks !!

//Does the header refer to a general template? JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

//The abstract read more like an introduction, rather than an indication of the work undertaken for the paper. 25 “Results show that there is an effective transfer of information and that the agent learns successfully to solve the game efficiently.” //This is too vague. What are the outcomes of the study?

We need to change the abstract according to all the changes that we have to do.

39 This information is used to make a gaming agent improves its operational performance using electroencephalography (EEG) signals as feedback of the performed task, obtained from an observational human critic. //There is a grammar issue. I don’t understand this sentence.

We modified the sentence.

54 RL should be defined in main body of the paper. Also the term ‘agent’ needs to be defined/clarified. How does it relate to the Game Manager from Figure 1?

The acronym was revised and we also verified all the other acronyms used in the text. The term agent was defined, and their relation with the Game Manager in Figure 1 was established.

17 col 2 Recently, this technique has seen a come-back. // this is not scientific language

This mistake has been corrected.

58 col 3 The precision of 25.125 is inappropriate P2, 46 col 2 What is “state information”?

The precision for this information was adjusted to two decimals for the age mean and standard deviation. We removed the ”state information” and replaced the wording of the phrase in order to clarify more clearly the information that we wanted to convey.

P3, 2 red normally indicates an error

We understand the Reviewers point, and we agree with him that red is not the right color to represent the completion of a task in the game. However, we do not want to change it this point because is not trivial the screen configuration that we used in the experiments that elicited the ErrP response, which may be affected by colors and shapes ?.

P3, 11 I assume that the “observational human critic” is the player/participant/subject/human observer. This role should be clarified and terminology used consistently.

Absolutely. I didn't decide if use OHC "observational human critic" or subject throughout all the text.

P3, 41 define MNE

Historically MNE stood for Miimun Error Estimate, a software package developed in Martinos Center of Harvard University. Now MNE is the name of the entire software platform to perform analysis of several type of brain signals like Magnetoencephalography and Electroencephalography.

P3, 56 Thus, each epoch is composed of a matrix 500 x 8. // add channels

Excellent.

P4, 32 Hence, following the iterative procedure based on Equation 1, the Q-Table is updated in each iteration. After the algorithm finishes iterating through all the training episodes, the Q-Table is stored to test the performance of the agent. // Will the game always terminate? How long does the game take? Does smooth progression toward finish affect the err potential?

If the gaming agent moves randomly on the board of Figure 2, it takes on average 100 steps to arrive to the final location on the grid. Each step, the movement direction is selected from the Q-Table once every 2 seconds, so on average it takes around 200 seconds to finish the game. After some training, the Q-Table could potentially end up in loops so if the steps count arrives to 200 the game is interrupted and it starts all over again (and in that case no reward is received).

We didn't test if the err potential was affected by the smooth progression toward the end. It's something interesting to test.

We added all this information in the paper.

P4 Fig 3. Is the chance score = 0.5?

Yes, it is. We added that information in the Figure caption.

P4 the labels referred to in Fig 5 should correlate with the test here, i.e., A= subject 1 etc.

The figure labels were modified to reflect which graph references which subject, in a more clearly way.

P4 referring to Fig 6; a comparison ROC curve with subject 1 would be more interesting

We added the ROC Curve for subject 1.

P5, 34 what is meant by “experiences”?

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

P5, 29 col 2 – remove Average steps per Q-Table legend.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

P5 50, col 2 The collected data show that ErrP signals can in fact be classified and used to train an agent effectively. // how has effectiveness been determined here?

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 6 Fig 10 the text is not legible, it should be improved.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 6, 40 “However, even though this implies that the agent misses frequently that an action taken is wrong, this is not hindering the overall performance and the agent is still learning.” // Your results show that this is subject dependent P6, 56 Results show that training a classifier with data of one subject, but

using it to classify the events of experiences of another subject does not lead to an improvement on the performance of the agent.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

// could a pre-trained generic classifier provide a better initial state, subsequently trained with observer data to converge more quickly? Are there differential err potential for up, down, left, right?

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

### *Reviewer 3 General Comments*

Comments to the Author The following work shows the usage of ErrPs for training a gaming agent using reinforcement learning. Authors attempt different conventional classification approaches, as well as intersubject classification.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

It is unclear to me the benefit of evaluating single subject to single subject offline classification accuracy. Please elaborate on this.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Since different subjects attained different performance, all combinations of subjects used for training and testing should be inspected for the 1:1 evaluation setup.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Maybe classification accuracy could be reported at different steps (depending on the amount of training data – gradual increasing the number of subjects)

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

What is the ratio of hit/no hit segments of action (epochs)?

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 3, column 2, lines 4-9. It is unclear at this stage what action from the starting point would generate an ErrP. moves- $\zeta$  move

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 3, column 2, lines 11-13. Has the MinMax Scaler been applied in any other BCI related study or elsewhere? I think including a reference would be useful.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 4, Figure 3. I would suggest adding another set of bars for the grand average classification scores.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 4, Figure 4. This figure is redundant and could be removed.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 5, Figure 5. The titles of the subplots would be more descriptive when replacing the letters with the corresponding subjects number.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 5, column 1, line 52, Not  $\zeta$ no

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 5, column1, line 55, accumulative- $\zeta$ cumulative

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 6, Figure 8. It is unclear to me what is the benefit of showing both A and B. Since the behavior is similar in both cases.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 6, Figure 10. This figure is, in my opinion, incomplete. What is the reason for showing this subset of subjects? I suggest showing the confusion matrices of all subjects and/or their average.

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

Page 6, column 2, line 40. show  $\zeta$  shows

This is a shameful mistake, we are terrible sorry. We thought this was going to be replaced by the platform automatically. Thanks for pointing out this error.

.....