The Board of Directors

German Research Foundation

Kennedyallee 40

53175 Bonn, Germany

xxx

02 October 2019

--•-,

German Research Foundation - 53170 Bonn, Germany

Personal/ Confidential

Professor Dr. Dr. h.c. Niels Birbaumer Eberhard Karls University of Tübingen

Institute for Medical Psychology and Behavioural Neurobiology

Silcher Road 5

72076 Tübingen Germany

GZ: 1-CWV-WI 29.05.07

Formal investigation procedure closed by decision of the Joint Committee on 19 September 2019

Dear Professor Birbaumer,

At its meeting on 19 September 2019, the Joint Committee of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) decided on the allegations of scientific misconduct made against you and established that in the publications

Chaudhary U, Xia, B, Silvoni, S, Cohen, L G, & Birbaumer, N (2017).

"Brain-computer inter-face-based communication in the completely

/ocked-in state". PLoS Bio/, 15(1), e1002593 (hereinafter referred to as PLoS 2017)

and

Chaudhary U, Pathak S, and Birbaumer N (2019) Response to: "Questioning the evidence for BCI-based communication in the complete locked in state". PLoS Bio/ 17(4): e3000063. (hereinafter referred to as PLoS 2019).

You have made false statements within the meaning of Section 11.1.a) of the DFG's Rules of Procedure for Dealing with Scientific Misconduct (VerfOwF). The Joint Committee considered this to be scientific misconduct and adopted the following measures within the meaning of Subclause 111.3.c) of the Constitution: Exclusion from eligibility to apply for funding for five years, non-use as reviewer for a period of five years and request to withdraw the incriminated publications. In addition, the Steering Committee has decided to reclaim from you the funds used for the incriminated publications, provided that these can be clearly allocated and quantified after an examination of the research line of the Reinhart Koselleck Project concerned.

Specifically, at the proposal of the Committee for the Investigation of Allegations of Scientific Misconduct, the Joint Committee found in three cases false information within the meaning of Clause 11.1.a) of the VerfOwF with regard to the publications PLoS 2017 and PLoS 2019:

Kommentar [F1]: This refers to false statements or misrepresentation of data

1. Incomplete video recordings and only summary statistics in PLoS 2017

The publication PLoS 2017 on page 18 states: "all sessions were videotaped and are available on request (...)".

According to your statement at the meeting of the Committee for the Investigation of Allegations of Scientific Misconduct on 9 July 2019, "feedback sessions" and "open question sessions" were recorded in principle. You admitted, however, that it had sometimes been forgotten to activate the video recorder or that there had been some technical problems, so that the recordings were at least incomplete and did not even exist for some questions. As a result, not everything that should have been recorded was recorded.

The members of the Committee pointed out that their examination of the publication PLoS 2017 and the data on which it was based had revealed that in 2014, in two patients (Patient F and Patient G), both the "feedback sessions" and the "open question sessions" had only been evaluated as sum statistics by the computer and that the individual questions and the answers given to them had not been broken down accordingly.

On the basis of the video recordings (if they had been completely available), the individual results could have been determined in spite of the summary statistics - both the questions as well as the answers given out loud by the computer (question/answer pairs) - in the aftermath, so that the facts claimed in PLoS 2017 on the concrete response behaviour of the patients could have been verified.

In your hearing you admitted that the missing video recordings had caused you "headaches". This, along with your statement that sometimes the activation of the VCR had been “forgotten” amounts to an admittance that you knew about the fact that the video recordings were incomplete. Since you also confirmed that you had discussed with Dr. Chaudhary both the fact of the summary statistics, which were initially only made, and how to deal with them, as well as the fact that some of the video recordings were missing, you knew that some of the individual results could not be given (contrary to what was stated in PLoS 2017) and that the video data was not available in full at the time. You thus acted at least grossly negligent with regard to the presentation of the results in PLoS 2017 and the underlying data.

On the basis of your statement, the Main Committee came to the conclusion that in PLoS 2017 you had provided an overall depth of data that is de facto non-existent and that the fact of false declarations within the meaning of Clause 11.1.a) of the VerfOwF had thus been realised.

Answer: For every session a video camera was taken and activated according to the study protocol. The videos were in principle taken for demonstration purposes and not for verification of the data, which was not necessary, since all relevant data were stored in the computer and the videos only showed that a yes or no answer was activated. In the early sessions a very old camera was used and the camera sometimes stopped recording. In other instances personal care of the patient required to turn off the camera and sometimes it was not immediately turned on again. In some instances the video camera was not recording at all. Thus, in fact only a proportion of the sessions had complete videotapes. We were not aware of this fact at the time of the submission of the article but found problems in the video documentation only in the course of the investigation of scientific misconduct. Thus the statement in the methods section that all sessions were videotaped is correct but it must be amended by stating that not all video sessions were complete. We have thus not made any false statement in the paper. Eine Falschaussage würde implizieren, dass der Umstand bei Einreichung des Artikels bekannt war; dies war jedoch nicht der Fall.

With respect to the summary statistics it must be noted that 6 sessions in patient G and 3 sessions in patient F – which were conducted at the beginning of the study - contained summary statistics. This was related to problems in the program which could not be fixed during a site visit with the patients. However, we have not reported anywhere in the methods section that a single trial analysis was conducted. This would also not have been necessary, since the ability to communicate was evaluated exclusively based on the percentage of correct yes-no answers. Thus we have not made any false statement in the paper.

In addition, BCI studies do not usually employ video recordings nor do they use single trial analyses. We therefore do not believe that the summary sessions are a methodological problem of the study. We computed the results of our study with and without the summary sessions and the sessions with single trial data available had an even better outcome (Summary sessions patient G. 17% above threshold; patient F: 33% above threshold; individual trial sessions: patient F: 75% above threshold; patient B: 100% above threshold; patient W: 50% above threshold). Thus taking out the summary sessions would greatly improve our percentage correct responses.

Aside from this, the exact questions are documented and the answers can be fully reconstructed from the brain data that were classified to activate the yes and no answers. Thus the data are completely determined in the documentation we provided. Since the videos are not a compulsory part of this type of set-up and the yes no answers can be reconstructed from the data that were uploaded, we do not believe that this invalidates our results.

We did not know about the quality of the video recordings or that the summary statistics might be a problem when we submitted the paper. The conversations about incomplete documentation refer to the time when we were investigated for scientific misconduct and carefully examined all our data. Thus we did not intentionally suppress information when submitting the paper and we reject the statement that we purposefully made any false declarations based on the facts provided above. In fact, the sessions without complete video documentation had worse results that the

sessions with complete video documentation (68.18 correct without video versus 70.63 correct with video.

We would like to point out that the whistleblower never raised this point, which was added to the investigation by the DFG committee and we object to the inclusion of this point in the investigation.

2 Exclusion of data from the publication PLoS 2017

In the publication PLoS 2017 on page 9 it says: "None of the sessions were eliminated... ".

In the course of clarifying the allegations, however, it became apparent that a large number of training sessions were excluded from publication in all patients, mostly for technical reasons, but once for unknown reasons, without this being disclosed in the publication (sufficiently comprehensible).

In the opinion of the Joint Committee, pre-defined criteria for the inclusion and exclusion of data by the authors were not sufficiently plausibly presented in the publication.

Moreover, the basic requirements for clinical studies have not been taken into account to the necessary extent, according to which all examinations of patients must be meticulously documented. In the opinion of the Steering Committee, this documentation obligation also includes indications of missing data or technical problems with data collection.

Answer: On page 9 in the discussion section of the paper (this is not the methods section!) we exclusively refer to the feedback sessions (N=21) not to the training sessions, since the determination of communication ability we refer to on page 9 was only based on the feedback sessions. We had to exclude 0 of 21 sessions.

The training sessions the reviewers refer to were not addressed in this statement on page 9 of the paper. When training sessions were excluded for technical reasons, this related to the fact that triggers were not properly transmitted and the data could not be analysed. The following sessions contained technical errors related to equipment malfunction in the training sessions: Patient F: 9/70 sessions, B: 2/48 sessions, G: 7/66 sessions, and patient W:1/21 sessions. This is not large amount given that the training was done under difficult conditions in the patients' home environment and not in the laboratory. Thus, also for the training session, no data that could be analysed were excluded from the publication. In addition, PLoS Biology guidelines ask to upload the data that were used to generate the results. The data with technical errors could not be analysed at all. And therefore could not be considered for the results.

As for the session that was eliminated without reason: we did not eliminate any session without reason and we cannot follow this argument.

Since this was not a clinical study but an exploratory study that examined these patients with these methods and this equipment for the first time, it was not possible to predefine all unforeseen circumstances in the home of these patients. Asking this would preclude any study that examined new scientific questions and uses new methods. It is clear from our paper that this was not a clinical study. Wir möchten auch darauf verweisen, dass es sich hier um eine im Koselleck-Programm der DFG geförderte Studie handelte, bei der es eine Voraussetzung ist, innovative und risikobehaftete Forschungsvorhaben zu realisieren. In an exploratory study it is simply not possible to anticipate every circumstance that could make it necessary to exclude data. As with every scientific study, sessions were excluded in which such serious technical errors occurred that they could not be evaluated. We do not know of any published study that would report technical

failures of the equipment if these sessions cannot be included in the data analysis (e.g. missing cable in an electrical stimulation study which leads to the stimulation not arriving and an invalid data collection). Furthermore, training sessions were excluded that did not meet the criteria for Yes-No differentiation (described in the paper on page 18). In addition, sessions in which patients needed medical attention were excluded (described in the paper on page 17). Again, to the best of our knowledge we documented the analyzable data that could not be included in the model building and we do not see how the inclusion of the data with technical failure would have changed anything in the results of our paper. We transmitted them to the DFG to be comprehensive but still do not think that they should have been uploaded with the data for PloS Biology. We of course uploaded the training sessions that were omitted from the model building stage since they could be analyzed.

In table 1 in PLoS 2017 a certain number (51) of training sessions - ",,(d) number of sessions averaged/patient F)" - was also specified, but these were partially excluded in the algorithm ("classifier") underlying the machine learning approach.

Answer: In none of the patients all training sessions were used to build the model but only sessions where the yes no differentiation exceeded the pre-specified threshold. This is presented in the methods section on page 18: "If the classification accuracies for at least three consecutive [a]training sessions with questions with known answers were greater than the chance-level threshold, a new model was generated using the relative change in O2Hb across three training sessions to give online feedback". The other training session did not contribute to the model building.

Later Figure S2 in PLoS 2019 again referred to these 51 training sessions.

Answer: The Figure S2 in the 2019 publication refers to patient B. We cannot understand this comment, which seems to be based on patient F.

Furthermore, two out of three visits of patient F (visits 1 and 2) could not be evaluated as part of the underlying PLoS 2017 study. The study was registered post hoc with clinicaltrials.gov (November 2016) with study start in June 2014 (both visits had already taken place before the start date).

Answer: Registration of an experimental study with patients was not required in 2014. By the time of the submission of the study in 2016 this became more common also for non-clinical studies that included patients and PLoS Biology asked to register the study retroactively, which was only possible 24 months back. As noted, this study was never conceptualized as a clinical study and the exclusion criterion does therefore not hold for the first two visits of patient F. This was an exploratory study with patients but not a clinical study. Even if we exclude the two visits, the results would not change.

We are also surprised that completely new points have been raised by the committee (exclusion of a session without reason, clinical study aspects), which were never discussed with us. This issue was also never raised by the whistleblower.

As an experienced senior professor, you should have known that, especially in clinical trials, selection criteria for the inclusion or exclusion of data must be defined in advance, and that data excluded from the publication must be stated, stating the circumstances justifying the exclusion. It should therefore have been obvious to you that you did not proceed lege artis with regard to the clinical study or the corresponding data collection/analysis/documentation.

Answer: This is not a clinical study and therefore these criteria do not hold. Registering with clinicaltrials.gov does not imply a clinical study, it only implies that patients were treated. It can also be an exploratory study. As noted above, this type of study was performed on these patients for the first time and is to be considered an exploratory study. The technical problems encountered in the clinical handling of the patients could not all be defined with foresight, since this type of measurement feedback was carried out for the first time. We would again like to point out that this study was part of a DFG-funded Reinhart Koselleck project, which requires to conduct "exceptionally innovative or higher-risk projects".

As a result, the Board has also affirmed a scientific misconduct in the form of false declarations for the aforementioned facts.

Answer: As noted above we strongly disagree with these conclusions based on a wrong evaluation of the facts of this study.

3. false declaration in PLoS 2019

The results shown in Fig. S2 in PLoS 2019 could not be replicated - so the reproach -

- using the so-called t-test on the basis of the data already published in the supplement to PLoS 2017.

With regard to this accusation, you have stated that an independent expert had also made recalculations with regard to Fig. S2 and had achieved even better results than you yourself.

In the opinion of the Joint Committee, the basic analysis of the data (sequence from averaging to GLM analysis) was methodologically/statistically incorrect.

In essence, however, the present procedure was not about verifying or falsifying your result or the result of the person providing the information. Rather, the task was to examine whether the respective information in the publication was correct. You yourself stated in your written statement of 25 June 2019 and in your hearing on 9 July 2019 that the results could not have been reproduced at all, as this was not possible with the help of the data uploaded to PLoS 2017 in the Supplement. Thus, in the opinion of the Main Committee, it has been admitted that the publication PLoS 2017 was at least accompanied by an insufficient data set.

Answer: The T-test and the GLM analysis are unrelated analyses and it seems that the committee has falsely assumed that we did these analyses in sequential order, but these were unrelated procedures used by us to characterize the data. But aside from this, the use of the proper test statistic is a matter of scientific discourse not a matter of scientific misconduct.

As for the T-test results presented in Figure 2 of the 2019 Supplement we would like to emphasize that the data underlying the t-test were uploaded on the server in 2017 and were available to the whistleblower or anybody else who wanted to replicate the data.

The data in Figure S2 of 2019 were unrelated to the data in Table S2 in the original 2017 publication, which referred to T-tests between the fNIRS and the EEG data and did not contain the time series data but individual accuracy values, which no t-Test data can be computed. The whistleblower stated that he used these data for a t-test. However, these data are only the data underlying the figure and cannot be used for the

calculation of a t-test. We stated to the DFG that in contrast to what the whistleblower said the t-test analyses could not have been done based on the data of the S2 Table in the 2017 article since PLoS Biology only asks to display the data represented in the Figure not the raw data. The whistleblower would have had to go to the uploaded data to calculate the t-Tests from time series presented there.

It is surprising that a wrong calculation of the whistleblower is now held against us.

On the basis of this, the Steering Committee also found that the publication PLoS 2019 was incorrect in the sense of section 11.1.a) of the VerfOwF.

Answer: The committee did not recognize that the whistleblower made a mistake in trying to analyze data that were unrelated to the 2019 figure rather than using the uploaded raw data. We therefore refute these allegations.

Finally, we would like to inform you that neither the Steering Committee nor the external experts have made any statements as to whether the communication you claim to have with CLIS patients works or not. The subject of the DFG procedure was and is only the evaluation of your scientific working methods measured against the standards of good scientific practice or scientific misconduct, not the verification/falsification of the research results themselves.

Answer: We would like to point out that the committee addressed multiple new points that were never brought up by the whistleblower and that should in our opinion not have been part of the investigation. The committee also failed to note that there was no evidence for invented or suppressed data and manuscripts or falsification of data nor wrong computation of data as originally stated by the whistleblower. The committee also did not bring in an expert for fNIRS and BCI research who could have documented that the computations of the whistleblower contain serious errors that cannot be blamed on us.

Finally, we can document that the points the DFG committee raised had no effect on the data. We therefore do not understand why the paper in PloS Biology should be retracted.

The DFG procedure is concluded with the decision of the Joint Committee. We will inform the Rector of the University of Tübingen and the whistleblowers about the decision of the Joint Committee.

Answer: We have documented above that there were errors in the evaluation of the committee. It is in our opinion not in accordance with good scientific practice and our legal system that no recourse to this committee decision is possible.

Sincerely,

Xxxx

(president)