

## LETTER TO THE EDITOR

### Reply: On assessing neurofeedback effects: should double-blind replace neurophysiological mechanisms?

Manuel Schabus<sup>1,2</sup>

1 Laboratory for Sleep, Cognition and Consciousness Research, Department of Psychology, University of Salzburg, Austria

2 Centre for Cognitive Neuroscience Salzburg (CCNS), University of Salzburg, Austria

Correspondence to: Manuel Schabus, PhD,  
University of Salzburg,  
Hellbrunnerstr. 34, 5020 Salzburg,  
Austria  
E-mail: Manuel.Schabus@sbg.ac.at

Sir,

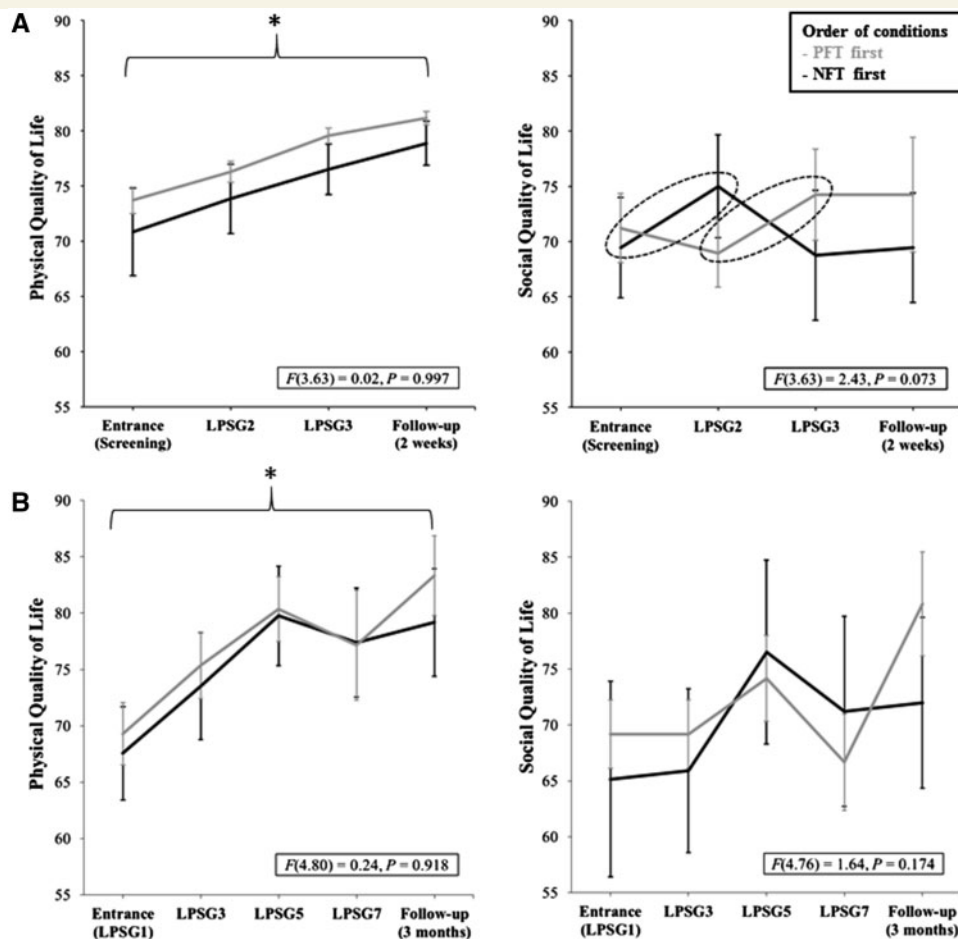
We highly appreciate an open discussion regarding the effects of neurofeedback training (NFT), wherefore we are happy to respond to the letter by Fovet and colleagues, 2017.

We agree that, strictly speaking, the results of our study (Schabus *et al.*, 2017) do not allow generalizing the negative findings reported in primary insomnia to other NFT applications. Yet, what is disturbing is the fact that even misperception insomniacs (i.e. participants without any objectively quantifiable sleep problems) show unaltered EEG activity [in the same 12–15 Hz range; sensorimotor rhythm (SMR)] minutes after NFT sessions ended. At least we would have expected a sustained effect in this subgroup because they should not suffer from severe learning impairments. Surprisingly, these findings contradicted earlier reports from our own laboratory (Hoedlmoser *et al.*, 2008; Schabus *et al.*, 2014), where we added a healthy and younger control group and verified that participants achieved similar SMR enhancements during NFT training blocks (14–28%) as found earlier (15–25%; Schabus *et al.*, 2014).

Fovet and colleagues question how exactly these discrepancies between earlier (positive) results of our group in healthy participants (Hoedlmoser *et al.*, 2008) and insomniacs (Schabus *et al.*, 2017) are explained. First of all, it is important to note that in all three studies, we used exactly the same NFT methodology. The only exception was that we extended the NFT from 8 × 3 min to 8 × 5 min blocks (including two ‘transfer blocks’ with no immediate feedback) in the last double-blind study (Schabus *et al.*,

2017). This was specifically done because we expected participants to have shallower learning curves due to the higher age [mean = 38.59, standard deviation (SD) = 11.18 in Schabus *et al.* (2017); mean = 34.83, SD = 10.60 in Schabus *et al.* (2014); mean = 23.63, SD = 2.69 in Hoedlmoser *et al.* (2008)] and the higher severity of insomnia [e.g. polysomnography (PSG)-derived wake after sleep onset 64.56 min in Schabus *et al.* (2017) versus 37.01 min in Schabus *et al.* (2014)]. So first of all, the slightly higher age may have rendered the observed effects in the latest study smaller than in the earlier two NFT studies. Second, we clearly highlight the non-optimal study design in our first insomnia study (2014), which was intended to serve as a comprehensive pilot test for the present and much more controlled double-blind study. Not only was that earlier study only single-blind but it also compared 10 blocks of NFT to five blocks of placebo-feedback (i.e. likewise randomized-frequency feedback). In an earlier study (*cf.* Fig. 6; Schabus *et al.*, 2014), we reported a placebo effect on subjective quality of life across the sessions, i.e. an effect that was independent of whether participants received placebo or real neurofeedback; reanalysing the subjective data, we indeed found evidence that patients may have felt more social support in that earlier single-blind study in the NFT condition. This should be alarming for any neurofeedback study not adopting a double-blind design (*cf.* Fig. 1) as this effect is unlikely limited to our NFT study.

Also, note that in this earlier study, we were unable to find an increase in sleep spindles or memory performance following NFT, contrary to what we had found in healthy



**Figure 1** Comparison of the subjective quality of life data from our earlier single-blind (Schabus *et al.*, 2014) and current double-blind (Schabus *et al.*, 2017) study. (A) Subjective quality of life effects (WHOQOL) are plotted for the sub-domain physical quality of life (including activities of daily living, energy and fatigue, pain and discomfort, sleep and rest or work capacity) and social quality of life (including personal relationships and perceived social support) for our earlier single-blind study. Note that physical quality of life changes over time but irrespective of placebo-feedback training (PFT) or NFT. Data for social quality of life indicate a trend towards increased perceived social support only between sessions with real NFT training [dashed circles; i.e. entrance to LPSG2 in NFT-first group ( $n = 12$ ); and LPSG2 to LPSG3 in the PFT-first group ( $n = 11$ )]. (B) The same subjective data for our double-blind study (Schabus *et al.*, 2017) do not indicate differences in perceived social support (for NFT versus PFT training). Yet we found again a non-specific increase in physical quality of life from study entrance (LPSG1) to the follow-up after 3 months. Note that here LPSGs 1 and 3 flank the first training block ( $12 \times$  NFT or PFT) and LPSGs 5 and 7 flank the second training block ( $12 \times$  NFT or PFT). PFT-first participants included 10 and NFT-first 12 participants (i.e. participants with WHOQOL values for all five time points).  $F$ -values depict the interaction between group (NFTfirst, starting with NFT, PFTfirst, starting the protocol with PFT) and time [entrance, learning nights/polysomnographies (LPSG), and Follow-up]. Error bars indicate  $\pm 1$  standard error.

young controls (Hoedlmoser *et al.*, 2008). What we did find was a linear relationship of SMR enhancement (i.e. an SMR amplitude change of NFT Session 2–3 to Session 9–10) during NFT and a respective (fast) spindle number change. Splitting the sample into NFT ‘responders/learners’ and ‘non-responders/non-learners’ as performed in many NFT studies seemed artificial and thus not justified to us. Indeed, splitting participants into ‘learners’ and ‘non-learners’ or ‘SMR increasers’ and ‘SMR decreasers’ (Reichert *et al.*, 2015; Mayer *et al.*, 2016) appeared to be a common practice in the field. However, this approach is circular as it confirms, in the worst case, no more than that for example SMR increasers or learners (as defined by the investigators)

indeed increase in SMR or slow cortical potential amplitude without adding any further knowledge.

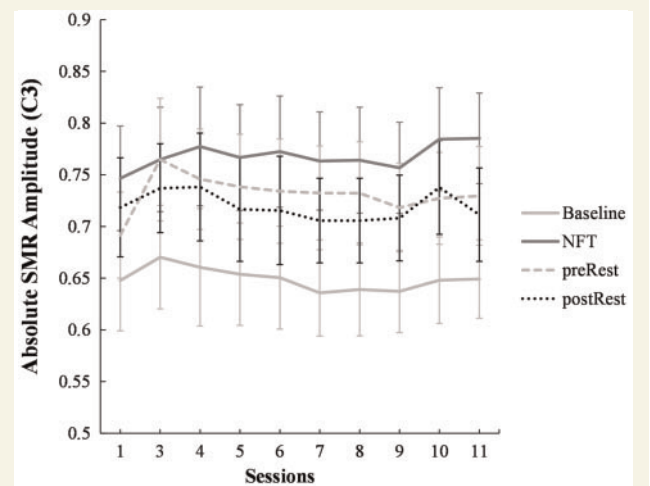
The authors continue to refer to a ‘well-controlled double-blind study’ by Kober *et al.* (2015) that contradicted our recent findings and crucially found an increase in relative SMR amplitude and declarative memory performance following NFT. Yet, we are not convinced by the findings presented in this study. For the SMR, Kober *et al.* only found within-session changes (together with a similar increase in a non-rewarded 21–35 Hz beta band) but essentially ‘no significant changes in absolute SMR power over the feedback training sessions’, which we had reported earlier (Hoedlmoser *et al.*, 2008). According to our understanding, only systematic

power changes over NFT sessions (i.e. multiple days) would show actual learning over time. For behavioural changes, the authors analysed 12 cognitive parameters of which three were significant in a pre- to post-session comparison (at  $P < 0.05$ ) in the NFT group versus only one in the control group. Simulating the data reveals that the only pre- to post-change that survives a correction for multiple comparisons ( $P < 0.005$ ) is the VVM2 construction 1 (Visual and Verbal Memory Test by Schelling and Schächtele, 2001). Importantly, however, participants' scores are not significantly different between the experimental and the control groups [ $t(18) \approx 0.024$ ;  $P \approx 0.980$ ]. Thus, we neither see evidence for the increase of SMR amplitude (over sessions) nor a difference in behavioural performance between the experimental and control groups.

We believe that in many NFT studies, one factor giving rise to misleading effects may be that the control groups are (if present at all) often a 'playback' NFT session, i.e. the NFT feedback another subject has received is just replayed to a 'control' participant. This kind of control appears extremely problematic as participants in the control group then learn over extended periods of time that they have no control over the feedback received. Essentially, manipulations of that kind may be used for learned helplessness but do not depict an adequate control condition for neurofeedback as it will likely induce negative training effects in the controls (against which NFT is tested thereafter). In order to circumvent this bias, we tried to carefully match the feedback received in our NFT and placebo-feedback conditions [1686 for placebo-feedback ( $SD = 208$ ) and 1652 trials for NFT ( $SD = 277$ )] with thresholds adapted after each 5-min run to stay in the range of 13–25 rewards.

Fovet *et al.* are also concerned that increasing the amount of rewards received within a run may have rendered the training too easy in the current protocol. We do not share this concern as the current approach leads to a reward about every 17s, and if participants exceeded 25 rewards within a 5-min run, we increased the threshold to be exceeded for the following 5 min feedback period to assure that the training remained challenging. Furthermore, as stated earlier, the changes in SMR from baseline are identical in the 2014 (15–25%) and 2017 (14–28%) studies; it is a misconception that the earlier changes were 115–125% as mentioned in Fovet *et al.*'s letter (100% was taken as baseline level, or 'zero'; cf. Fig. 2 caption, Schabus *et al.*, 2014).

We completely agree with Fovet and colleagues that a better understanding of neurophysiological mechanisms involved in various neurofeedback protocols is needed. However, screening the NFT literature we see very few studies actually addressing questions of that kind and we believe that highly controlled, i.e. double-blind, studies with combined EEG-functional MRI might be especially suited to demonstrate systematic brain changes related to various neurofeedback protocols convincingly.



**Figure 2 Absolute SMR amplitude across the neurofeedback sessions.** Note that SMR amplitude (on trained electrode C3) is (i) higher for the NFT training period ('NFT') as compared to the baseline before ('Baseline') each training trial [main effect for Condition,  $F(1.21) = 115.98$ ,  $P < 0.001$ ]; as well as (ii) higher for the NFT training period as compared to the eyes-open resting condition before the training started on that day ['preRest'; main effect for Condition,  $F(1.20) = 6.61$ ,  $P = 0.018$ ]. We here pooled all insomnia and misperception insomniacs with sufficient data points at each session ( $n = 22$  for Baseline to NFT period;  $n = 21$  for preRest to NFT period). To derive absolute SMR amplitude, artefact-corrected EEG data were transferred to the frequency domain by applying the FFT (fast Fourier transform) to 1-s segments. Spectral line values were calculated using one half of the spectrum and are expressed in  $\mu V/2$ . To reduce artefacts caused by potential signal discontinuities at the segment borders, segments were tapered using a Hanning window (window length 10%). Finally, a periodic variance correction was applied to account for the reduction in signal strength induced by the windowing function. Error bars indicate  $\pm 1$  standard error.

Indeed, there are some studies that persuasively demonstrate that certain frequencies can be upregulated across training sessions (i.e. across days and not within a training session) in young healthy individuals (Hoedlmoser *et al.*, 2008; Philippens and Vanwersch, 2010; Zoefel *et al.*, 2011; Arns *et al.*, 2014) using NFT. Yet, at the same time there are several studies only showing changes within runs (i.e. within a training session) but not across sessions (Dempster and Vernon, 2009; Kober *et al.*, 2015; Reichert *et al.*, 2015). Although our data indicate that we could increase SMR amplitude 'significantly from its spontaneous value', i.e. in the NFT training period as compared to the respective NFT baselines; as well as compared to the eyes-open rest condition (cf. Fig. 2). We do not see EEG changes from before to after NFT or increased (absolute) SMR amplitude across sessions [e.g. Session 1 to Session 11:  $t(26) = -1.34$ ,  $P = 0.19$ ; Fig. 2].

We believe that our data (Schabus *et al.*, 2017) therefore support the view that although EEG changes are readily observed within an NFT session, systematic changes in oscillatory brain activity over time are generally hard to achieve in the elderly or in patients with some kind of

learning impairments. Yet without detectable changes in brain activity, which are at least maintained over a number of hours (here, a rest condition flanking each NFT), days and weeks (here, across the NFT sessions), it is hard to imagine how neurofeedback can lead to consistent amelioration of various disorders and symptoms as often claimed. Most critically, the vast majority of neurofeedback studies do not present EEG data at all (Cortooos *et al.*, 2010; Gruzelier, 2014; Reiner *et al.*, 2014), therefore it is impossible to evaluate the credibility of their outcome beyond subjective reports of symptom change or some behavioural tests, which could be explained by a placebo effect as well.

In conclusion, our publication was designed to demonstrate that increasing 12–15 Hz SMR activity and sleep spindles can improve sleep and ameliorate primary insomnia symptoms. Therefore, we did not intend to refute neurofeedback research altogether. Yet, given our negative findings and reviewing a broad range of published neurofeedback studies, we believe that we highlight key issues that need to be addressed in the whole field. Importantly, addressing these issues and concerns will increase the credibility of the field, which we see as a goal worth striving for. Specifically, we are not aware of any similarly well-controlled study as the one published in *Brain* earlier this year (Schabus *et al.*, 2017) for any NFT protocol, study sample, or patient group. Given the fact that many of the reviewed studies in the field solely rely on subjective data, have no or insufficient control groups, and seem to build upon the *a priori* assumption that NFT has to have an effect, the scientific foundation of neurofeedback still appears to stand on rather shaky ground. It is also likely that publication bias (Kuhberger *et al.*, 2014) is further supporting the idea of neurofeedback as a universally efficacious non-pharmacological treatment in the field simply because negative findings may be seriously under-represented in the literature.

For the reasons we outline above, we strongly disagree with the authors that double-blind studies are premature for a field whose most significant contributions date back to the 70s and 80s (Serman *et al.*, 1970; Hauri, 1981; Hauri *et al.*, 1982). Most importantly, we once walked into the trap ourselves, that is, we jumped to some premature conclusions underestimating single-blind limitations and placebo effects associated with NFT in our earlier pilot study (Schabus *et al.*, 2014).

As stated previously in our empirical contribution, we sincerely welcome further rigorously controlled neurofeedback studies that critically address different kinds of NFT protocols and study samples. Unfortunately at present, an overwhelming amount of NFT studies are not withstanding high scientific standards, and are more harmful than helpful for the field and it almost seems that the NFT community is still relying on laurels gained years ago. Nevertheless, the

general rationale to directly target those brain oscillations that are altered in a specific disorder or associated with an improvement in performance is appealing. Therefore, we sincerely hope that the field will find ways to convince the scientific community that neurofeedback has indeed to be considered as a non-pharmacological alternative for various disorders and/or peak performance training.

## Funding

Research was supported by the FWF research grant P-21154-B18 and Y777 from the Austrian Science Fund.

## References

- Arns M, Feddema I, Kenemans JL. Differential effects of theta/beta and SMR neurofeedback in ADHD on sleep onset latency. *Front Hum Neurosci* 2014; 8:1019.
- Cortooos A, De Valck E, Arns M, Breteler MHM, Cluydts R. An exploratory study on the effects of tele-neurofeedback and tele-biofeedback on objective and subjective sleep in patients with primary insomnia. *Appl Psychophysiol Biofeedback* 2010; 35:125–34.
- Dempster T, Vernon D. Identifying indices of learning for alpha neurofeedback training. *Appl Psychophysiol Biofeedback* 2009; 34:309–18.
- Fovet T, Micoulaud-Franchi J-A, Vialatte F-B, Lotte F, Daudet C, Batail J-M, et al. On assessing neurofeedback effects: should double-blind replace neurophysiological mechanisms?. *Brain* 2017; 140: e63.
- Gruzelier JH. Differential effects on mood of 12-15 (SMR) and 15-18 (beta1) Hz neurofeedback. *Int J Psychophysiol* 2014; 93:112–15.
- Hauri P. Treating psychophysiological insomnia with biofeedback. *Arch Gen Psychiatry* 1981; 38:752–8.
- Hauri PJ, Percy L, Hellekson C, Hartmann E, Russ D. The treatment of psychophysiological insomnia with biofeedback: a replication study. *Biofeedback Self Regul* 1982; 7:223–35.
- Hoedlmoser K, Pecherstorfer T, Gruber G, Anderer P, Doppelmayr M, Klimesch W, et al. Instrumental conditioning of human sensorimotor rhythm (12-15Hz) and its impact on sleep as well as declarative learning. *Sleep* 2008; 31:1401–18.
- Kober SE, Witte M, Stangl M, Valjamae A, Neuper C, Wood G. Shutting down sensorimotor interference unblocks the networks for stimulus processing: an SMR neurofeedback training study. *Clin Neurophysiol* 2015; 126:82–95.
- Kuhberger A, Fritz A, Scherndl T. Publication bias in psychology: a diagnosis based on the correlation between effect size and sample size. *PLoS One* 2014; 9: e105825.
- Mayer K, Blume F, Wyckoff SN, Brokmeier LL, Strehl U. Neurofeedback of slow cortical potentials as a treatment for adults with attention deficit-hyperactivity disorder. *Clin Neurophysiol* 2016; 127:1374–86.
- Philippens IH, Vanwersch RA. Neurofeedback training on sensorimotor rhythm in marmoset monkeys. *Neuroreport* 2010; 21:328–32.
- Reichert JL, Kober SE, Neuper C, Wood G. Resting-state sensorimotor rhythm (SMR) power predicts the ability to up-regulate SMR in an EEG-instrumental conditioning paradigm. *Clin Neurophysiol* 2015; 126:2068–77.
- Reiner M, Rozenfurt R, Barnea A. Better than sleep: theta neurofeedback training accelerates memory consolidation. *Biol Psychol* 2014; 95:45–53.
- Schabus M, Griessenberger H, Gnjecza MT, Heib DPJ, Wislowska M, Hoedlmoser K. Better than sham? A double-blind placebo-controlled

- neurofeedback study in primary insomnia. *Brain* 2017; 140:1041–52.
- Schabus M, Heib DPJ, Lechinger J, Griessenberger H, Klimesch W, Pawlizki A, et al. Enhancing sleep quality and memory in insomnia using instrumental sensorimotor rhythm conditioning. *Biol Psychol* 2014; 95:126–34.
- Schelling D, Schächtele B. VVM – Verbaler und Visueller Merkfähigkeitstest. Göttingen: Hogrefe; 2001.
- Sterman MB, Howe RC, Macdonald LR. Facilitation of spindle-burst sleep by conditioning of electroencephalographic activity while awake. *Science* 1970; 167:1146–8.
- Zoefel B, Huster RJ, Herrmann CS. Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance. *Neuroimage* 2011; 54:1427–31.