# Statistical Analysis Plan (SAP)
## - Final Analysis -

## A single-centre, assessor-blinded, randomized-controlled feasibility study of near infrared spectroscopy neuro feedback for binge-eating disorder

# NIRSBED

Clinical trial protocol: Final2.0/ 18.01.2018

**Coordinating investigator / Investigators**

Prof. Dr. Anja Hilbert / Dr Ricarda Schmidt
University of Leipzig Medical Centre
IFB AdiposityDiseases, Behavioural Medicine Research Unit
Department of Psychosomatic Medicine and Psychotherapy
Semmelweisstrasse 10
04103 Leipzig

**Biometry / Project management**

Dr habil. David Petroff /
Dr Christiane Prettin
University of Leipzig
Clinical Trial Centre Leipzig - KKS
Härtelstraße 16-18
04107 Leipzig

Date of version: 21.05.2021
State of version: 1.0 final

We hereby approve the Statistical Analysis Plan in its final version:

Coordinating
Investigator

25.05.2021

Date

Signature

Prof. Dr Dipl.-Psych. Anja Hilbert

Biostatistician

27.05.21

Date

Signature

Dr rer. nat. habil. David Petroff

**Table of contents**

# 1 Introduction

The purpose of this document is to provide a detailed elaboration of the statistical analysis described in the protocol, including detailed procedures for the confirmatory analysis of the primary and secondary endpoints and other variables.

The SAP is based on the planned analysis specification as written in the study protocol, section 7 "Biometric Aspects of the Study". All data will be analysed and reported in accordance with a recently published checklist (CRED-nf checklist; Ros et al., 2020). The table can be found in appendix 6.4.

The software package R Version 3.6.1 or higher will be used for statistical analyses. To reproduce online analysis, Turbo-Satori (TSI) 1.0.0. (BrainInnovation B.V., Maastricht, the Netherlands; Lührs & Göbel, 2017) will be used. Offline analyses of fNIRS data will be carried out with MATLAB R2018b using the Brain AnalyzIR Toolbox (Santosa et al. 2018); see Item 4e of the CRED-nf checklist).

Database variables are denoted by `this font and colour`.

# 2 Endpoints and further variables

## 2.1 Primary endpoint

The number of objective binge-eating episodes (OBEs) over the past 28 days at post-assessment (T1, `visit_label=T1` for the intervention groups) or after the waiting period (`visit_label=T0WL` for the waitlist control group; Item 6a of the CRED-nf checklist). OBEs are derived from the well-established Eating Disorder Examination interview (`oepism1`) and will be compared between groups, correcting for the baseline value (`visit_label=T0`). The trial protocol stipulated that the number of OBEs over the past 14 days would be analysed. Based on the standard EDE instrument and the accepted endpoint, the value over the past 28 days will be used (Item 6a of the CRED-nf checklist).

## 2.2 Secondary endpoints (Item 6a of the CRED-nf checklist)

Unless otherwise indicated, the scores from questionnaires are calculated using the "qscorer" package from Norbert Köhler (see appendix for code). Comparisons between all three groups are foreseen at T0 (baseline) and T1/T0WL. After T0WL, the WL control group received an fNIRS intervention. Hence, T2 can only be used to estimate changes within a group and between the two intervention groups.

a) Feasibility (between T0-T1)

- Recruitment: a graphical depiction of recruitment will be provided along with the mean recruitment rate

- Attrition: The number of treatment sessions attended in the intervention groups (based on `exdat`) will be provided by randomization arm and taking dropout due to SARS-CoV-2 into account

- Assessment completion: Valid OBEs at T1 will be counted as "assessment completion"

- Compliance: Therapist's evaluation at T1 on personal effort, general motivation, and treatment success (`eff3, moti4, succ5`), scale between 0-5

- Patients' program evaluation: Patient's response in the last session attended to question 3 (`PRE31, PRE32, PRE33`): "How helpful was the neurofeedback training until now in your opinion (a) in general, (b) as it pertains to your eating behaviour,(c) in order to relax"?, scale between 0-6

b) Eating disorder psychopathology

- Number of OBEs (T0, T2) defined as for the primary endpoint

- Abstinence from objective binge eating (T1-T2) meaning `oepism1==0`

- Remission (T1-T2), (remission means `bedfull= bedlfrq = bedlimdu = nein`), partial remission (T1-T2), defined as change from `bedfull` to `bedlfrq` or `bedlimdu` – only those with `bedfull` at T0 will be included in the endpoint analysis of partial remission. All data will be included in descriptive statistics.

- Eating disorder psychopathology (EDE-Q) (T0-T2): the global score and the four subscale-scores (restraint, weight concern, shape concern, eating concern) will be used

c) Neuropsychological outcomes (T0-T2)

- Cognitive flexibility: Trail Making Test (TMT-B), completion time in s, age- and sex-specific percentile rank

- Planning: Tower of London, number of correct responses, age- and sex-specific percentile rank

- Inhibition: Stop Signal, stop signal reaction time in ms, age- and sex-specific percentile rank; Go/NoGo, cued Go/NoGo, number of commission errors, age- and sex-specific percentile rank

- Decision making: Cards and Lottery Task, number of advantageous decisions

- Food-specific attention: Food-specific Dot-Probe, attentional bias score in ms

- Food-specific approach/avoidance: Food-specific Approach-Avoidance Task, difference score in ms per category (high-calorie, low-calorie, neutral-boring, neutral-pleasant)

- Food-specific inhibition: Food-specific Stop Signal Task, stop signal reaction time in ms

d) Brain-activity-related outcomes

- fNIRS and fNIRS delayed treatment group, group-wise: Condition-(neurofeedback/transfer), block(first block/second block) and task-wise (regulation/mirror/rest) oxygenated, deoxygenated, and total haemoglobin values per session for the 2 individual fNIRS channels that were selected for training and for regions of interest.

- EEG group only (T0, T1): Eyes-closed (3min) and eyes-open (3min) resting-state and food-specific activity (90s) for theta, alpha, and beta band in µV

- EEG group only: mean beta activity baseline and feedback in µV per session noted from TheraPrax device, amount of successful trials and % success in transfer trial per session

e) Weight management-related behaviours (T0-T2)

- Self-efficacy (GSES): Sum score

- Impulsivity (BIS/BAS): Two scales: Behavioral Inhibition System (BIS) mean score and Behavioral Activity System (BAS) mean score

- Cognitive emotion regulation (CERQ): Sum scores for nine different conceptual scales: 1) Self-blame, 2) Acceptance, 3) Rumination, 4) Positive refocusing, 5) Planning, 6) Positive reappraisal, 7) Putting into perspective, 8) Catastrophizing and 9) Blaming others

- Emotion regulation (DERS): Global sum score. The six sub-scales will only be reported if there is a significant difference between randomization arms for the global score.

- Food Craving Questionnaire-Trait (FCO-T-r): Sum score

f) Mental health (T0-T2)

- Depressive symptoms (PHQ-D): Sum score

- Anxiety disorder symptoms (PHQ-A): Sum score

- Quality of life (SF-12): Physical and mental sum scores

g) Physical health (T0-T2)

- Body mass index (BMI, `weight/height^2*10^4`), kg/m²

- Hip and waist circumference, cm


## 2.3 Further variables

Control variables at baseline (T0) are used to characterize the patients and may be relevant for outcomes, so that particular attention is warranted to ensure that random differences between the groups are not large.

- Vigilance: Perception and attention functions battery (WAFA), log-transformed mean reaction times in ms, measure of dispersion (variability in performance in intrinsic and phasic vigilance) in ms

- Intelligence: Raven's Coloured Progressive Matrices, number of correct responses, age- and sex-specific percentile rank

- Working memory: *N*-back task, number of correct responses, age- and sex-specific percentile rank

Variables that may be relevant for session-wise and overall neurofeedback outcomes assessed in each session:

- Helpfulness of the training: general (PRE31), related to eating behaviour (PRE32), to relax (PRE33), scale between 0-6

- Questions about strategies: Attempts to apply strategies in daily life during the past seven days (PRE41), success in applying the strategies used during training (PRE42), success in applying these strategies related to eating behaviour (PRE431) and to relax (PRE432)

- PHQ-15: sum score (variables PRE51 – PRE515)

- Strategies that were used for the session (POST21), reported in an exemplary fashion (Item 3c of the CRED-nf checklist)

- Success in regulating brain activity during the past session (POST22), scale between 0-6

### 2.3.1 Baseline characteristics

The following characteristics of the trial population at baseline will be calculated / reported descriptively by randomization arm.

- Sex (sex)

- Age (difference between randodat and birth, for the latter, only month and year are available and the first of the month will be used)

- Level of education as a binary variable ("A-levels" or not, gradu==Abitur/Fachabitur)

- BMI (weight/height^2*10^4), and according to the stratification variable

- Weight status (normal weight, overweight, obese, according to BMI)

- Participation in a weight-loss programme (stratification variable), yes/no

- Waist circumference (waist) stratified by sex, cm

- Hypertension (dis1suf), yes/no, and medication for hypertension (dis1med)

- Type II diabetes (dis5suf), yes/no, and medication for diabetes (dis5med)

- Previous comorbid psychiatric disorders (PSYDIS), yes/no

- Previous psychotherapeutic or psychiatric treatment (THR), yes/no

- Participation in previous self-help groups (SPPGRP), yes/no

- Lowest (LOWGHT) and highest adult weight (HIWGHT), kg

- Number of medically controlled (MDCNTR) and non-medically controlled (NMDCNTR) previous weight loss attempts, scale between never - ≥ 5 times

- Type of motivation to participate in the study (weight loss [MNREAS]/voluntary control of brain activity [MNREAS1])

- Motivation to change eating behaviour (MTVN), readiness to maintain changes in eating behaviour in the long-term (WILL), confidence in maintaining changes in eating behaviour in the long-term (CONFI), scale between 1-10

This list may be modified for the publication.


## 3   General analysis definitions


### 3.1   Analysis populations

Analysis of the primary and secondary outcomes are based on the full analysis set (FAS). Note that patients fulfilling the criterion for abstinence at T0 will not be included in the analysis for this endpoint. Further, the analysis of the primary outcome will be performed in the per-protocol set (PPS) to evaluate the treatment effect for patients with good protocol adherence, if FAS and PPS comprise a sufficiently different set of patients. Analysis of the fNIRS data will be performed in the fNIRS data set.

The FAS includes all randomized patients unless purely technical issues precluded the fNIRS measurements. In particular, three patients had to be replaced because of the aforementioned technical issues with fNIRS. Patients who could not complete the intervention due to the SARS-CoV-2 situation will be retained in the FAS if they completed at least 5 sessions and provided OBE data at T1. In control patients, those who could not attend T0WL due to SARS-CoV-2 will not be retained in the FAS.

The PPS includes all patients of the FAS without serious violations of the study protocol. The following protocol violations are classified as serious:

- Violation of at least one inclusion or exclusion criterion

- Treatment was not conducted as mandated, that is
  - Temporal delay of last attended treatment session by 4 weeks relative to the planned date[1], `exdat`, (based on the date of the first treatment session). Note that consideration of a maximum of two weeks' delay in each session as originally foreseen was not considered feasible.
  - Absence of more than 6 of the planned 12 treatment sessions (based on `meet`)
- Use of forbidden medication
- Psychological treatment during study participation (was not the case, otherwise it would be documented in the "comments")

The fNIRS/EEG data set includes all patients from the fNIRS/EEG group of the FAS who provided data for more than a single neurofeedback session. Participants with fewer sessions are excluded from the analyses, because the definition of neurofeedback *training* does not apply for these participants and there is no information on between-session variability.

## 3.2   Subgroups

No subgroups will be analysed.

# 4   Planned analysis

A flowchart according to the CONSORT statement will describe all patients randomized to the trial detailing withdrawals, drop-outs and inclusion in the analysis sets defined above. It will also provide details regarding treatment and assessment completion.

Standard methods of descriptive statistics will be used always indicating the number of valid values; frequencies, mean (standard deviation), minimum, maximum and/or quartiles (median [25%; 75%]), depending on the scale and distribution within the study population and/or per arm will be provided. Summary statistic will be reasonably rounded to avoid pseudo-precision. Reliability (Cronbach's α) will be calculated for each psychometric scale at T0.

For the explorative neurophysiological analyses, only completer analyses will be considered and in patients who have provided neurophysiological data on at least two sessions. There the primary focus is on understanding  neurophysiological processes in those that complete the training sessions and its potential effect on outcome.

## 4.1   Primary endpoint

The primary endpoint, the number of OBEs at post-assessment or post-waiting, will be analysed using ANCOVA with OBEs as the dependent variable, OBEs at T0 as a covariate as well as the stratification criteria (`strabmi, strasex, strawei`). The code for this test will have the structure

```
anova(lm(OBE_FU ~ OBE_T0 + STRABMI + STRASEX + STRAWEI + ARMS)).
```

The variables OBE_FU and OBE_T0 will be shifted by the mean of OBE at T0 so that the intercept in the model reflects change in time. A closed-testing procedure will be used, meaning

---

[1] The planned dates are as follows: session 1/2 (week 1), session 3/4 (week 2), session 5/6 (week 3), session 7/8 (week 4), session 9 (week 5), session 10 (week 6), session 11 (week 7), session 12 (week 8),

that the initial comparison will be between the two intervention groups combined and the waiting group and the three pair-wise tests (fNIRS vs. wait control group, EEG vs. wait control group and fNIRS vs. EEG group) will be performed without adjustment if the global test is significant. If the global test is not significant, estimates and confidence intervals will still be provided for the contrasts in OBE, but no formal statistical test will be carried out. Estimates will be taken from the linear model and a Wald confidence interval will be used. Moreover, secondarily, the analogous analysis with effect size will be performed.

Missing data will be treated with multiple imputation using the R package "mice" (van Buuren and Groothuis-Oudshoorn 2011). The imputation variables will include the answer to the question "In how many of the situations in which you ate too did you feel you had lost control of your eating behaviour"[2] (`pre72`). Sex, BMI and the stratification variable participation in a weight-loss programme will also be used. Fifty imputations will be performed.

To test the robustness of the results, the primary analysis will be repeated without imputation and pessimistic and optimistic scenarios will be considered.

## 4.2 Secondary endpoints

### 4.2.1 Analysis of binary variables

Abstinence from objective binge eating and remission as defined above will be compared between randomization arms at T1/T0WL using an exact Fisher test and presenting the proportion of patients abstinent/in remission per arm with a Wilson confidence interval. In addition, a logistic regression with the stratification variables and categories for the number of binge-eating episodes at baseline (0, 1–3, 4–7, 8–13, >13) as covariates will be performed. If the model converges, odds ratios for the difference between arms will be taken from the coefficients and confidence intervals using a profiling method. If there are problems with convergence, then the number of categories in the covariates will be reduced and, if need be, covariates will be removed.

The abstinence from binge eating and remission from BED data at T2 for the intervention groups and end of treatment for the control group will be presented, but no formal statistical test will be performed.

### 4.2.2 Analysis of continuous variables

Mean and SD will be provided at all available points in time for all valid values and after imputation. Imputation variables will include all data from the same "category" of secondary endpoints along with age and sex. If the imputation methods should fail to converge, then imputation variables will be successively removed starting with those having the weakest absolute value for linear correlation.

Variables at T1/T0WL will be analysed with linear models having the same structure as for the primary endpoint, but will include all three randomization arms from the outset (i.e. will not make use of the closed testing principle). Here too the variables will be shifted as with the primary endpoint.

Data for T0, T1 and T2 will be analysed for the intervention groups in a mixed model in which the stratification variables, time and group will be treated as fixed effects including the interaction between group and time. The patient ID will be treated as a random term. The model will be fit using restricted maximum likelihood (REML) and an unstructured variance-covariance matrix.

If multiple sub-scales of a questionnaire are tested formally, then Bonferroni-Holm correction for multiplicity will be used.

---

[2] The original German reads „In wie vielen dieser Situationen, in denen Sie zu viel gegessen haben, hatten Sie das Gefühl, die Kontrolle über Ihr Essverhalten verloren zu haben (während des Essens)?"

## 4.3 Analysis of fNIRS data

Certain aspects of the analysis of fNIRS data have not yet been clarified so that we now present only a temporary plan to be updated in the next version of the SAP.

The pre-processing of the fNIRS data follows standard procedures and is described below (appendix 6.11). After converting the raw data to haemoglobin concentrations, these are analysed in a general linear model (GLM) where they are treated as the dependent variable with the predicted time course as the independent variable (design matrix). The model estimates a weighting for this predicted time course (referred to here as β).

The ability of the GLM to model the data may be very limited. In such cases, the type I error is not likely to be inflated, but the type II error can be extremely large, i.e. regardless of the effect size, the model will have very low power. If the mean explained variance of the GLM at baseline is below 15%, then the model will be deemed inadequate and will not be used.

In the standard pre-processing pipeline, outliers are identified on the task and section level and removed. The number of such data (participants) will be reported by session and task (Item 3e of the CRED-nf checklist) and in sensitivity analyses their effect will be studied.

### 4.3.1 Regulation success based on the feedback signal (CRED-nf 5a)

A graphical depiction for summary statistics of the raw data for each session in the regulation condition will be provided for the following variables

- percentage success defined to be the number of sample points with online oxygenated haemoglobin values were smaller than the predecessor divided by the total number of sampling points (roughly $10^5$). This is a surrogate for the proportion of time for which the patient experienced "success" though a maximum/minimum picture size could not be exceeded and each step was so small that it could not be perceived.

- minimum picture size (12 per patient) as one measure of success for the feedback

- mean picture size (12 per patient) as a second measure of success for the feedback

Transfer condition

- a graphical representation as with the regulation condition will be provided

Note that after session 6, the selection of pictures was changed to avoid habituation effects.

### 4.3.2 Within-session and between-session comparisons (CRED-nf 5b/5c)

The β-values for the two particular channels of each patients' individual ROIs will be analysed. Since the predicted time course in the first and second half of the regulation block and also between blocks is identical, a comparison of β amounts to a comparison of the optimal weighting of this predicted time course. As noted above, this can be difficult to interpret: e.g. an increase from 0.8 to 1 could indicate that the response followed the expectation both times with an increase in amplitude from 0.8 to 1, but it is also conceivable that there was good response to the stimulus the first time and none whatsoever with the higher value of β.

- β will be displayed graphically for the first and second half of the average of all sessions

- The two β values per session will be averaged and displayed graphically per sessions

- β will be analysed after standardization in a mixed linear model with the coefficients as the dependent variable. For the within-session analysis, the first and second half of the regulation condition will be estimated (fixed effect) and session (linear fixed effect) and patient as a random effect for the intercept. For comparisons between the tasks regulation and mirror, a mixed model with condition as the fixed effect of interest, session and first/second half as a nuisance parameter and patients as a random term will be used. These analyses will be repeated for the transfer vs mirror condition.

### 4.3.3 Correlation between regulation and outcomes for patients who provided data for at least two feedback sessions (CRED-nf 6b)

A linear model will be used with the number of OBEs at post-assessment as the dependent variable, the value at T0 as covariate and mean picture size across all sessions as the variable of interest.

## 4.4 Analysis of EEG data

The pre-processing of the EEG data follows standard procedures and is described below (see appendix 6.12). The resulting data available to us at pre and post feedback are mean power in each of three frequency bands (θ, α and β) per patient, and task. Moreover, during the pre and post visits as well as each session, mean power for the subset of the β band is available for each patient, condition and task that was used for feedback.

### 4.4.1 Regulation success based on the feedback signal (CRED-nf 5a)

The proportion of successful feedback and transfer trials, defined as trials with mean subset of the beta band values below the baseline value will be reported per session. Whereas in the case of NIRS, data were available on the amount of time in which the participant was successful per session, here only a single mean power value is available.

### 4.4.2 Within-session and between-session comparisons (CRED-nf 5b/5c)

Mean beta band activity in µV for baseline, feedback and transfer trial will be displayed graphically per session.

For within-session variability, two mixed effects models (one for feedback and one for transfer) will be used. The dependent variable is the mean power in the β band and the fixed effects are the (constant) baseline mean power in the β band and the number of the trial, treated as a numerical variable (i.e. assuming linear behaviour over time). The patient is the random term.

To assess the success of the feedback relative β band power will be shown graphically for resting state eyes open, eyes closed, and food cue presentation at pre and post treatment. Paired t-test of the logarithms of these values will be used for formal testing with Bonferroni-Holm correction for the three states.

For comparison with a "control" condition, the difference of β band power on a logarithmic scale (i.e. logarithm of the ratio) from feedback to its baseline will be compared to the difference from transfer to its baseline. These differences will be treated in a mixed model with condition (feedback vs transfer) and session number fixed effects along with their interaction and patient as a random variable. The interaction term will quantify a potential learning effect, i.e. that the feedback in the later sessions is no longer necessary and the difference between the conditions gets smaller.

### 4.4.3 Correlation between regulation and outcomes for patients who provided data for at least two feedback sessions (CRED-nf 6b)

A linear model will be used with the number of OBEs at post-assessment as the dependent variable, the value at T0 as covariate and mean proportion of successful trials across all sessions as the variable of interest.

## 4.5 Interim analyses

No interim analyses were planned or performed.

# 5 Safety

Adverse events (AEs) and Serious Adverse Events (SAEs) as well as PHQ-15 were collected during the therapy sessions and will be listed by therapy arm. For PHQ-15, the raw data will be presented in graphical form per randomization arm and point in time.

# 6 Appendix

## 6.1 Abbreviations

AE          Adverse event

BMI         Body mass index

CONSORT     Consolidated Standards of Reporting Trials

OBE         Objective binge-eating episode

SAE         Serious adverse event

## 6.2 Code for EDE-Q

The EDE-Q instrument was introduced by Fairburn and Beglin (1994) and the scoring of the German version follows Hilbert and Tuschen-Caffier (2016).

```r
scoring_edeq <- function(data, items = 1:22, keep = TRUE, nvalid = list(
  restraint = 3,
  eating_concern = 3,
  weight_concern = 3,
  shape_concern = 4
), digits = NULL) {
  if (min(data[, items], na.rm = T) < 0) {
    stop("Minimum possible value for EDE-Q items is 0")
  } else if (max(data[, items], na.rm = T) > 6) {
    stop("Maximum possible value for EDE-Q items is 6")
  }
  # check for number of specified items
  if (length(items) != 22) {
    stop("Number of items must be 22!")
  }
  items.restraint <- items[1:5]
  items.eating_concern <- items[c(7,9,13,14,15)]
  items.weight_concern <- items[c(8,12,16,18,19)]
  items.shape_concern <- items[c(6,8,10,11,17,20:22)]
  data <- data %>%
    mutate(
      nvalid.edeq.restraint = rowSums(!is.na(select(., all_of(items.restraint)))),
      score.edeq.restraint = rowMeans(select(., all_of(items.restraint)), na.rm = TRUE),
      score.edeq.restraint = ifelse(nvalid.edeq.restraint >= nvalid[['restraint']],
score.edeq.restraint, NA),
      nvalid.edeq.eating_concern = rowSums(!is.na(select(., all_of(items.eating_concern)))),
      score.edeq.eating_concern = rowMeans(select(., all_of(items.eating_concern)), na.rm =
TRUE),
      score.edeq.eating_concern = ifelse(nvalid.edeq.eating_concern >=
nvalid[['eating_concern']], score.edeq.eating_concern, NA),
      nvalid.edeq.weight_concern = rowSums(!is.na(select(., all_of(items.weight_concern)))),
      score.edeq.weight_concern = rowMeans(select(., all_of(items.weight_concern)), na.rm =
TRUE),
      score.edeq.weight_concern = ifelse(nvalid.edeq.weight_concern >=
nvalid[['weight_concern']], score.edeq.weight_concern, NA),
      nvalid.edeq.shape_concern = rowSums(!is.na(select(., all_of(items.shape_concern)))),
      score.edeq.shape_concern = rowMeans(select(., all_of(items.shape_concern)), na.rm =
TRUE),
      score.edeq.shape_concern = ifelse(nvalid.edeq.shape_concern >= nvalid[['shape_concern']],
score.edeq.shape_concern, NA),
      score.edeq.sum = rowMeans(cbind(score.edeq.restraint, score.edeq.eating_concern,
score.edeq.weight_concern, score.edeq.shape_concern),na.rm=TRUE),
      score.edeq.sum = ifelse(rowSums(is.na(cbind(score.edeq.restraint,
score.edeq.eating_concern, score.edeq.weight_concern, score.edeq.shape_concern)))<=1,
score.edeq.sum, NA)
```

```
  )
  # Keep single items and nvalid variables
  if (keep == FALSE) {
    data <- data %>% select(-all_of(items), -starts_with('nvalid.'))
  } else {
    data <- data
  }
  # Rounding
  if (is.numeric(digits) == TRUE) {
    data <- data %>% mutate_at(vars(starts_with('score')), list(~ round(., digits)))
  } else {
    data <- data
  }
  data
}
```

## 6.3   Code for determining the GSES score

The source for determining the score is (Schwarzer 1999)

```
function (data, items = 1:10, keep = TRUE, nvalid = 7, digits = NULL)
{
    library(dplyr, warn.conflicts = FALSE)
    if (min(data[, items], na.rm = T) < 1) {
        stop("Minimum possible value for GSES items is 1")
    }
    else if (max(data[, items], na.rm = T) > 4) {
        stop("Maximum possible value for GSES items is 4")
    }
    if (length(items) != 10) {
        stop("Number of items must be 10!")
    }
    items <- items
    data <- data %>%
      mutate(nvalid.gses = rowSums(!is.na(select(., items))),
        mean.temp = round(rowSums(select(., items), na.rm = TRUE)/nvalid.gses)) %>%
      mutate_at(vars(items), list(~ifelse(is.na(.), mean.temp, .))) %>%
      mutate(score.temp = rowSums(select(., items), na.rm = TRUE),
        score.gses = ifelse(nvalid.gses >= nvalid, score.temp, NA)) %>%
      select(-mean.temp, -score.temp)
    data
    if (keep == FALSE) {
        data <- data %>% select(-items, -nvalid.gses)
    }
    else {
        data <- data
    }
    if (is.numeric(digits) == TRUE) {
        data <- data %>% mutate_at(vars(score.gses), list(~round(.,digits)))
    }
    else {
        data <- data
    }
    data
}
```

## 6.4   Code for determining the BIS/BAS score

The questionnaire was originally derived  by Carver and White (1994) and is described on the website https://local.psy.miami.edu/people/faculty/ccarver/availbale-self-report-instruments/bisbas-scales/. The German version works with means (instead of sums) and only two subscales (Strobel et al. 2001), though more recent articles suggest other choices (Müller et al. 2013), but conclude that due to limited evidence, one retain the two factors.

```
scoring_bisbas_german <- function(data, items = 1:24, keep = TRUE,
                     nvalid = list(
                         bis = 6,
                         bas = 10
                     ),
```

```
                                digits = NULL) {

  if (min(data[, items], na.rm = T) < 1) {
    stop("Minimum possible value for BISBAS items is 1")
  } else if (max(data[, items], na.rm = T) > 4) {
    stop("Maximum possible value for BISBAS items is 4")
  }
  # check for number of specified items
  if (length(items) != 24) {
    stop("Number of items must be 24!")
  }
  items.rev <- items[c(1,3:21,23,24)]
  items.bis <- items[c(2,8,13,16,19,22,24)]
  items.bas <- items[c(3,4,5,7,9,10,12,14,15,18,20,21,23)]
  data <- data %>%
    mutate_at(vars(all_of(items.rev)), list(~ 5 - .)) %>%
    mutate(
      nvalid.bisbas.bis = rowSums(!is.na(select(., all_of(items.bis)))),
      nvalid.bisbas.bas = rowSums(!is.na(select(., all_of(items.bas)))),
      score.bisbas.bis = rowMeans(select(., all_of(items.bis)), na.rm = TRUE),
      score.bisbas.bis = ifelse(nvalid.bisbas.bis >= nvalid[['bis']], score.bisbas.bis, NA),
      score.bisbas.bas = rowMeans(select(., all_of(items.bas)), na.rm = TRUE),
      score.bisbas.bas = ifelse(nvalid.bisbas.bas >= nvalid[['bas']], score.bisbas.bas, NA),
    )
  # keep/discard original items and nvalid vars
  if (keep == FALSE) {
    data <- data %>% select(-all_of(items), -starts_with('nvalid.'))
  } else {
    data <- data
  }
  # Rounding
  if (is.numeric(digits) == TRUE) {
    data <- data %>% mutate_at(vars(starts_with('score')), list(~ round(., digits)))
  } else {
    data <- data
  }

  data
}
```

## 6.5  Code for determining the CERQ score

The questionnaire is derived and described in Garnefski et al. (2001) and a manual provided in Garnefski and Kraaij, V. Spinhoven, P. (2002). The shortened German version is described in Görgen et al. (2015). Originally, one missing item out of four per sub-score was allowed, but since the German version has only three items per sub-score, we do not permit any missing items.

```
scoring_cerq_german <- function(data, items = 1:27, keep = TRUE,
                                digits = NULL) {

  if (min(data[, items], na.rm = T) < 1) {
    stop("Minimum possible value for BISBAS items is 1")
  } else if (max(data[, items], na.rm = T) > 5) {
    stop("Maximum possible value for BISBAS items is 5")
  }
  # check for number of specified items
  if (length(items) != 27) {
    stop("Number of items must be 27!")
  }
  items.self_blame <- items[0:2*9+1]
  items.acceptance <- items[0:2*9+2]
  items.rumination <- items[0:2*9+3]
  items.positive_refocusing <- items[0:2*9+4]
  items.refocus_planning <- items[0:2*9+5]
  items.reappraisal <- items[0:2*9+6]
  items.perspective <- items[0:2*9+7]
  items.catastrophizing <- items[0:2*9+8]
  items.blaming_others <- items[0:2*9+9]
  data <- data %>%
    mutate(
      nvalid.cerq.self_blame = rowSums(!is.na(select(., all_of(items.self_blame)))),
```

```
      score.cerq.self_blame = rowSums(select(., all_of(items.self_blame)), na.rm = FALSE),
      nvalid.cerq.acceptance = rowSums(!is.na(select(., all_of(items.acceptance)))),
      score.cerq.acceptance = rowSums(select(., all_of(items.acceptance)), na.rm = FALSE),
      nvalid.cerq.rumination = rowSums(!is.na(select(., all_of(items.rumination)))),
      score.cerq.rumination = rowSums(select(., all_of(items.rumination)), na.rm = FALSE),
      nvalid.cerq.positive_refocusing = rowSums(!is.na(select(.,
all_of(items.positive_refocusing)))),
      score.cerq.positive_refocusing = rowSums(select(., all_of(items.positive_refocusing)),
na.rm = FALSE),
      nvalid.cerq.refocus_planning = rowSums(!is.na(select(.,
all_of(items.refocus_planning)))),
      score.cerq.refocus_planning = rowSums(select(., all_of(items.refocus_planning)), na.rm =
FALSE),
      nvalid.cerq.reappraisal = rowSums(!is.na(select(., all_of(items.reappraisal)))),
      score.cerq.reappraisal = rowSums(select(., all_of(items.reappraisal)), na.rm = FALSE),
      nvalid.cerq.perspective = rowSums(!is.na(select(., all_of(items.perspective)))),
      score.cerq.perspective = rowSums(select(., all_of(items.perspective)), na.rm = FALSE),
      nvalid.cerq.catastrophizing = rowSums(!is.na(select(., all_of(items.catastrophizing)))),
      score.cerq.catastrophizing = rowSums(select(., all_of(items.catastrophizing)), na.rm =
FALSE),
      nvalid.cerq.blaming_others = rowSums(!is.na(select(., all_of(items.blaming_others)))),
      score.cerq.blaming_others = rowSums(select(., all_of(items.blaming_others)), na.rm =
FALSE)
    )
  # keep/discard original items and nvalid vars
  if (keep == FALSE) {
    data <- data %>% select(-all_of(items), -starts_with('nvalid.'))
  } else {
    data <- data
  }
  # Rounding
  if (is.numeric(digits) == TRUE) {
    data <- data %>% mutate_at(vars(starts_with('score')), list(~ round(., digits)))
  } else {
    data <- data
  }

  data
}
```

## 6.6   Code for determining the DERS score

The questionnaire is derived and described in Gratz and Roemer (2004), but for the correct order of variables and re-poling, please see

```
scoring_ders <- function(data, items = 1:36, keep = TRUE,
                       nvalid = list(
                           nonacceptance = 5,
                           goals = 4,
                           impulse = 5,
                           awareness = 5,
                           strategies = 6,
                           clarity = 4
                       ),
                       digits = NULL) {

  if (min(data[, items], na.rm = T) < 1) {
    stop("Minimum possible value for BISBAS items is 1")
  } else if (max(data[, items], na.rm = T) > 5) {
    stop("Maximum possible value for BISBAS items is 5")
  }
  # check for number of specified items
  if (length(items) != 36) {
    stop("Number of items must be 36!")
  }
  items.rev <- items[c(1, 2, 6, 7, 8, 10, 17, 20, 22, 24, 34)]
  items.nonacceptance <- items[c(11, 12, 21, 23, 25, 29)]
  items.goals <- items[c(13, 18, 20, 26, 33)]
  items.impulse <- items[c(3, 14, 19, 24, 27, 32)]
  items.awareness <- items[c(2, 6, 8, 10, 17, 34)]
  items.strategies <- items[c(15, 16, 22, 28, 30, 31, 35, 36)]
  items.clarity <- items[c(1, 4, 5, 7, 9)]
  data <- data %>%
    mutate_at(vars(all_of(items.rev)), list(~ 6 - .)) %>%
    mutate(
```

```
        nvalid.ders.nonacceptance = rowSums(!is.na(select(., all_of(items.nonacceptance)))),
        nvalid.ders.goals = rowSums(!is.na(select(., all_of(items.goals)))),
        nvalid.ders.impulse = rowSums(!is.na(select(., all_of(items.impulse)))),
        nvalid.ders.awareness = rowSums(!is.na(select(., all_of(items.awareness)))),
        nvalid.ders.strategies = rowSums(!is.na(select(., all_of(items.strategies)))),
        nvalid.ders.clarity = rowSums(!is.na(select(., all_of(items.clarity)))),
        score.ders.nonacceptance = 6*rowMeans(select(., all_of(items.nonacceptance)), na.rm =
TRUE),
        score.ders.nonacceptance = ifelse(nvalid.ders.nonacceptance >= nvalid[['nonacceptance']],
score.ders.nonacceptance, NA),
        score.ders.goals = 5*rowMeans(select(., all_of(items.goals)), na.rm = TRUE),
        score.ders.goals = ifelse(nvalid.ders.goals >= nvalid[['goals']], score.ders.goals, NA),
        score.ders.impulse = 6*rowMeans(select(., all_of(items.impulse)), na.rm = TRUE),
        score.ders.impulse = ifelse(nvalid.ders.impulse >= nvalid[['impulse']],
score.ders.impulse, NA),
        score.ders.awareness = 6*rowMeans(select(., all_of(items.awareness)), na.rm = TRUE),
        score.ders.awareness = ifelse(nvalid.ders.awareness >= nvalid[['awareness']],
score.ders.awareness, NA),
        score.ders.strategies = 8*rowMeans(select(., all_of(items.strategies)), na.rm = TRUE),
        score.ders.strategies = ifelse(nvalid.ders.strategies >= nvalid[['strategies']],
score.ders.strategies, NA),
        score.ders.clarity = 5*rowMeans(select(., all_of(items.clarity)), na.rm = TRUE),
        score.ders.clarity = ifelse(nvalid.ders.clarity >= nvalid[['clarity']],
score.ders.clarity, NA),
      )
  # keep/discard original items and nvalid vars
  if (keep == FALSE) {
    data <- data %>% select(-all_of(items), -starts_with('nvalid.'))
  } else {
    data <- data
  }
  # Rounding
  if (is.numeric(digits) == TRUE) {
    data <- data %>% mutate_at(vars(starts_with('score')), list(~ round(., digits)))
  } else {
    data <- data
  }

  data
}
```

## 6.7   Code for determining the FCQ-T-r score

The score is introduced in Meule et al. (2014) and the cut-off in Meule (2018).

```
scoring_fcqtr <- function(data, items = 1:15, keep = TRUE,
                          nvalid = list(
                              sum = 12
                          ),
                          digits = NULL) {

  if (min(data[, items], na.rm = T) < 1) {
    stop("Minimum possible value for BISBAS items is 1")
  } else if (max(data[, items], na.rm = T) > 6) {
    stop("Maximum possible value for BISBAS items is 6")
  }
  # check for number of specified items
  if (length(items) != 15) {
    stop("Number of items must be 15!")
  }
  items.sum <- items[1:15]
  data <- data %>%
    mutate(
      nvalid.fcqtr.sum = rowSums(!is.na(select(., all_of(items.sum)))),
      score.fcqtr.sum = 15*rowMeans(select(., all_of(items.sum)), na.rm = TRUE),
      score.fcqtr.sum = ifelse(nvalid.fcqtr.sum >= nvalid[['sum']], score.fcqtr.sum, NA),
      cutoff.fcqtr.sum = case_when(
        score.fcqtr.sum <50 ~ "no food addiction",
        score.fcqtr.sum >=50 ~ "food addiction"
      ),
      cutoff.fcqtr.sum = factor(cutoff.fcqtr.sum, levels=c("no food addiction", "food
addiction"))
    )
  # keep/discard original items and nvalid vars
  if (keep == FALSE) {
    data <- data %>% select(-all_of(items), -starts_with('nvalid.'))
```

```
  } else {
    data <- data
  }
  # Rounding
  if (is.numeric(digits) == TRUE) {
    data <- data %>% mutate_at(vars(starts_with('score')), list(~ round(., digits)))
  } else {
    data <- data
  }

  data
}
```

## 6.8    Code for determining the PHQ-D score

The score is discussed in the following articles (Spitzer et al. 1999; Kroenke et al. 2010)

```
function (data, items = 1:9, keep = TRUE, nvalid = 7, digits = NULL)
{
    library(dplyr, warn.conflicts = FALSE)
    if (min(data[, items], na.rm = T) < 0) {
        stop("Minimum possible value for PHQ-9 items is 0")
    }
    else if (max(data[, items], na.rm = T) > 3) {
        stop("Maximum possible value for PHQ-9 items is 3")
    }
    if (length(items) != 9) {
        stop("Number of items must be 9!")
    }
    items <- items
    data <- data %>%
        mutate(nvalid.phq9 = rowSums(!is.na(select(.,items))),
          mean.temp = rowSums(select(., items), na.rm = TRUE)/nvalid.phq9) %>%
        mutate_at(vars(!!!items), list(~ifelse(is.na(.), mean.temp, .))) %>%
        mutate(score.temp = rowSums(select(., items), na.rm = TRUE),
          score.phq9 = ifelse(nvalid.phq9 >= nvalid,
          score.temp, NA), cutoff.phq9 = case_when(score.phq9 >= 20 ~ "Severe",
            score.phq9 >= 15 ~ "Moderately Severe", score.phq9 >= 10 ~ "Moderate",
            score.phq9 >= 5 ~  "Mild", score.phq9 < 5 ~ "Minimal"),
          cutoff.phq9 = factor(cutoff.phq9, levels = c("Minimal", "Mild", "Moderate",
            "Moderately Severe", "Severe")))%>%
        select(-mean.temp, -score.temp)
      if (keep == FALSE) {
        data <- data %>% select(-items, -nvalid.phq9)
      }
      else {
        data <- data
      }
      if (is.numeric(digits) == TRUE) {
        data <- data %>% mutate_at(vars(score.phq9), list(~round(.,
            digits)))
    }
    else {
        data <- data
    }
    data
}
```

## 6.9    Code for determining the PHQ-A score

The PHQ-A (also referred to as GAD-7, general anxiety disorder) score is derived from the PHQ questionnaire and was introduced in Spitzer et al. (2006) and an algorithm for missing values can be found in Teymoori et al. (2020).

```
scoring_gad7 <- function(data, items = 1:7, keep = TRUE, nvalid = 6, digits = NULL) {
  if (min(data[, items], na.rm = T) < 0) {
    stop("Minimum possible value for PHQ-9 items is 0")
  } else if (max(data[, items], na.rm = T) > 3) {
    stop("Maximum possible value for PHQ-9 items is 3")
  }
```

```
  # check for number of specified items
  if (length(items) != 7) {
    stop("Number of items must be 7!")
  }
  items <- items
  data <- data %>%
    mutate(
      nvalid.gad7 = rowSums(!is.na(select(., items))),
      score.gad7 = 7*rowMeans(select(., items), na.rm = TRUE),
      score.gad7 = ifelse(nvalid.gad7 >= nvalid, score.gad7, NA),
      cutoff.gad7 = case_when(
        score.gad7 >= 15 ~ "Severe",
        score.gad7 >= 10 ~ "Moderate",
        score.gad7 >= 5 ~ "Mild",
        score.gad7 < 5 ~ "Minimal"
      ),
      cutoff.gad7 = factor(cutoff.gad7, levels = c("Minimal", "Mild", "Moderate", "Severe"))
    )
  # Keep single items and nvalid variables
  if (keep == FALSE) {
    data <- data %>% select(-items, -nvalid.gad7)
  } else {
    data <- data
  }
  # Rounding
  if (is.numeric(digits) == TRUE) {
    data <- data %>% mutate_at(vars(score.gad7), list(~ round(., digits)))
  } else {
    data <- data
  }
  data
}
```

## 6.10 Code for determining the SF-12 scores

The following code follows the standard described in (Stieglitz 1999).

```
scoring_sf12 <- function(data, items = 1:12, keep = TRUE, digits = NULL) {

  if (min(data[, items], na.rm = T) < 0) {
    stop("Minimum possible value for SF-12 items is 0")
  } else if (max(data[, items], na.rm = T) > 6) {
    stop("Maximum possible value for SF-12 items is 6")
  }
  # check for number of specified items
  if (length(items) != 12) {
    stop("Number of items must be 12!")
  }
  items.rev6 <- items[c(1,8)]
  items.rev7 <- items[c(9,10)]
  data <- data %>%
    mutate_at(vars(all_of(items.rev6)), list(~ 6 - .)) %>%
    mutate_at(vars(all_of(items.rev7)), list(~ 7 - .))
  physical <- 56.57706 +
    Recode(data[,items[1]], "1= -8.37399; 2= -5.56461; 3= -3.02396; 4= -1.31872; 5=0") +
    Recode(data[,items[2]], "1= -7.23216; 2= -3.45555; 3=0") +
    Recode(data[,items[3]], "1= -6.24397; 2= -2.73557; 3=0") +
    Recode(data[,items[4]], "1= -4.61617; 0=0") +
    Recode(data[,items[5]], "1= -5.51747; 0=0") +
    Recode(data[,items[6]], "1=  3.04365; 0=0") +
    Recode(data[,items[7]], "1=  2.32091; 0=0") +
    Recode(data[,items[8]], "1= -11.25544;2= -8.38063; 3= -6.50522; 4= -3.80130; 5=0") +
    Recode(data[,items[9]], "1=  3.46638; 2=  2.90426; 3=  2.37241; 4=  1.36689; 5=  0.66514;
6=0") +
    Recode(data[,items[10]],"1= -2.44706; 2= -2.02168; 3= -1.61850; 4= -1.14387; 5= -0.42251;
6=0") +
    Recode(data[,items[11]],"1=  4.61446; 2=  3.41593; 3=  2.34247; 4=  1.28044; 5=  0.41188;
6=0") +
    Recode(data[,items[12]],"1= -0.33682; 2= -0.94342; 3= -0.18043; 4=  0.11038; 5=0")
  mental <- 60.75781 +
    Recode(data[,items[1]], "1= -1.71175; 2= -0.16891; 3=  0.03482; 4= -0.06064; 5=0") +
    Recode(data[,items[2]], "1=  3.93115; 2=  1.8684; 3=0") +
```

```
    Recode(data[,items[3]], "1=  2.68282; 2=  1.43103; 3=0") +
    Recode(data[,items[4]], "1=  1.44060; 0=0") +
    Recode(data[,items[5]], "1=  1.66968; 0=0") +
    Recode(data[,items[6]], "1= -6.82672; 0=0") +
    Recode(data[,items[7]], "1= -5.69921; 0=0") +
    Recode(data[,items[8]], "1=  1.48619; 2=  1.76691; 3=  1.49384; 4=  0.90384; 5=0") +
    Recode(data[,items[9]], "1= -10.19085;2= -7.92717; 3= -6.31121; 4= -4.09842; 5= -1.94949;
6=0") +
    Recode(data[,items[10]],"1= -6.02409; 2= -4.88962; 3= -3.29805; 4= -1.65178; 5= -0.92057;
6=0") +
    Recode(data[,items[11]],"1= -16.15395;2=-10.77911; 3= -8.09914; 4= -4.59055; 5= -1.95934;
6=0") +
    Recode(data[,items[12]],"1= -6.29724; 2= -8.26066; 3= -5.63286; 4= -3.13896; 5=0")
  # Keep single items
  if (keep == FALSE) {
    data <- data %>% select(-items)
  } else {
    data <- data
  }
  # Rounding
  if (is.numeric(digits) == TRUE) {
    data$score.sf12.physical <- round(physical,digits)
    data$score.sf12.mental <- round(mental,digits)
  } else {
    data$score.sf12.physical <- physical
    data$score.sf12.mental <- mental
  }
  data
}
```

## 6.11 CRED-nf checklist

| 1 Domain | Item # | Checklist item | Planned analysis reported in section |
|---|---|---|---|
| **Pre-experiment** | | | |
| | 1a | Pre-register experimental protocol and planned analyses | AO |
| | 1b | Justify sample size | SP |
| **Control groups** | | | |
| | 2a | Employ control group(s) or control condition(s) | AO |
| | 2b | When leveraging experimental designs where a double-blind is possible, use a double-blind | NA |
| | 2c | Blind those who rate the outcomes, and when possible, the statisticians involved | SP |
| | 2d | Examine to what extent participants and experimenters remain blinded | NA |
| | 2e | In clinical efficacy studies, employ a standard-of-care intervention group as a benchmark for improvement | SP |
| **Control measures** | | | |
| | 3a | Collect data on psychosocial factors | 2.3.1., A |
| | 3b | Report whether participants were provided with a strategy | AO |
| | 3c | Report the strategies participants used | AO |
| | 3d | Report methods used for online-data processing and artifact correction | 6.11.3., A |
| | 3e | Report condition and group effects for artifacts | 6.11.2., A |
| **Feedback specifications** | | | |
| | 4a | Report how the online-feature extraction was defined | 6.11.1., A |

| | | | | |
|---|---|---|---|---|
| | 4b | Report and justify the reinforcement schedule | 6.11.3., 6.12.1., A |
| | 4c | Report the feedback modality and content | 6.11.3., A |
| | 4d | Collect and report all brain activity variable(s) and/or contrasts used for feedback, as displayed to experimental participants | 4.3.2., 5.3.2., A |
| | 4e | Report the hardware and software used | 6.11.1., 6.11.3., 6.12.1., A |
| **Outcome measures** | | | |
| Brain | 5a | Report neurofeedback regulation success based on the feedback signal | 4.3.1., 4.4.1., A |
| | 5b | Plot within-session and between-session regulation blocks of feedback variable(s), as well as pre-to-post resting baselines or contrasts | 4.3.2., 4.4.2., A |
| | 5c | Statistically compare the experimental condition/group to the control condition(s)/group(s) (not only each group to baseline measures) | 4.3.2., 4.4.2., A |
| Behaviour | 6a | Include measures of clinical or behavioural significance, defined a priori, and describe whether they were reached | 4.1., A |
| | 6b | Run correlational analyses between regulation success and behavioural outcomes | 4.3.3., 4.4.3., A |
| **Data storage** | | | |
| | 7a | Upload all materials, analysis scripts, code, and raw data used for analyses, as well as final values, to an open access data repository, when feasible | NA |

*Note.* Darker shaded boxes represent *Essential* checklist items; lightly shaded boxes represent *Encouraged* checklist items. Article only: the respective item will be covered in the article, but does not apply to the statistical analysis plan; A: article, AO: article only, NA: not applicable, SP: see study protocol.

## 6.11. fNIRS Data Preprocessing

### 6.11.1. Region of Interest Selection

Prior to the first NF session, participants underwent a Go/NoGo and a passive viewing task that served as a functional localizer to delineate individual brain regions implicated in cognitive control.

A general linear model (GLM) with the standard hemodynamic response function and two regressors corresponding to the Go/NoGo and NoGo Passive viewing condition, respectively, was applied. The resulting *t*-values of beta weights for the difference between conditions (Go/NoGo – NoGoPassive Viewing) were used to determine the primary location for the fNIRS NF process: the two adjacent channels that showed the highest *t*-values in the same direction with the same sign were selected.

### 6.11.2. Offline Analysis of fNIRS data (Sarah Rösch)

Offline analyses of fNIRS data were calculated with MATLAB R2020a using the Brain AnalyzIR Toolbox (Santosa et al. 2018); see Item 4e of the CRED-nf checklist). The following parts of the default pipeline were used:

**Transformations**

Individual differences in head circumference have been shown to have a negligible effect on how the probe is positioned over the cortical region of interest for each participant and will therefore not be considered (Okamoto et al. 2004). The baseline of raw data was trimmed to 30s before and after each NF trial. Data were converted to changes in optical density values over time using the toolbox's *OpticalDensity* function. The function used in the toolbox is given by $\Delta OD(t) = -\log(I(t)/I0)$ where $I(t)$ is the intensity of the signal recorded and I0 is the mean reference signal intensity at baseline. The optical density values were converted to concentration changes using the modified Beer-Lambert law (via the toolbox's function *BeerLambertLaw)* under consideration of an age-dependent differential path length factor that is validated for frontotemporal regions (Scholkmann & Wolf 2013) and a partial path length factor. The age that is used for the calculation of this age-dependent differential path length factor was derived from the age at the first NF session.

**First Level Analysis**

A GLM was calculated for each channel, each half (first/second half of NF), each condition (feedback/transfer/mirror) and each session of each participant, modelling event-related concentration changes with a canonical hemodynamic response via the toolbox's function *GLM* (leaving the type at the default option AR-IRLS). The general linear equation was solved with an iteratively auto-regressive whitened, weighted least-squares method (AR-IRLS) (Barker et al. 2013). This model applies an $n^{th}$-order AR filter derived via an Akaike model-order selection to whiten both sides of the GLM expression. All source-detector pairs within a file are solved concurrently yielding a full covariance model of the noise. The AR-IRLS method accounts for the expected presence of serially correlated errors (i.e., extracerebral components) in the optical signals due to cardiac pulsation and respiration and has been shown to have a good control of type-I error (Huppert 2016). It is not recommended to filter data if using the AR-IRLS GLM algorithm (Huppert 2016). To ensure data quality, all data were visually inspected before and after applying the AR-IRLS method.

**Outlier Removal**

The toolbox offers the opportunity to automatically remove participants based on the leverage of the data in the *nirs.modules.RemoveOutlierSubjects*.

### 6.11.3. Online Analysis of fNIRS Data and Visual Feedback Signal

Turbo-Satori (TSI) 1.0.0. (BrainInnovation B.V., Maastricht, the Netherlands; Lührs & Goebel 2017; Item 4e) was used for real-time preprocessing. NIRStar 15.2. was connected to the NIRScout system via a universal serial bus cable and to TSI via a TCP/IP Ethernet connection.

The default settings of TSI were implemented as following: A GLM with ordinary least squares was fitted to the data. Raw wavelength data were detrended with a second-order moving average low-pass filter (cut-off 0.4Hz) to remove high-frequency noise (Item 3d of the CRED-nf checklist). These raw wavelength data were online converted to values of oxygenated and deoxygenated haemoglobin based on the modified Beer-Lambert law. In line with previous fNIRS NF studies, the feedback was based on the oxygenated haemoglobin signals. The NF signal was computed from the signal of the selected target channels by subtracting the baseline, which was set to the initial 20s of the experiment, from the currently scaled oxygenated haemoglobin value (Item 4a of the CRED-nf checklist).

**Visual Feedback Signal**

The online feedback was visually displayed on the screen as an individually appetitive food picture via a python built-in function. The size of the picture visualized the neural activity level in the chosen region of interest channels continuously (Item 4b and Item 4c of the CRED-nf checklist). Precisely, the oxygenated hemoglobin value for the region of interest (i.e., the average of the two channels that were selected as region of interest) for a given time point was compared to the oxygenated hemoglobin value for the region of interest at the preceding time point. The goal in the NF task was to increase the oxygenated hemoglobin value. Thus, if the oxygenated hemoglobin value exceeded the previous value, the picture size was reduced by the constant factor 0.01, if the oxygenated hemoglobin value was less than the previous value, the picture size was increased by the constant factor 0.01. Each new picture (of a total of $n = 12$ pictures) started with a picture size value of 1. The maximum picture size was defined by 1, the minimum picture size by 0.1. For these values, the picture size remained constant if participants decreased (for the maximum picture size) or increased (for the minimum picture size) their oxygenated hemoglobin value further. Nevertheless, these trials still count as success trials.

In the transfer condition, participants did not get real-time feedback. Instead, they were displayed the picture size they had reached by the final 2s of the experiment.

As outcome variable for the feedback condition, the minimum picture size, the mean picture size, and the percentage of successful trials were derived. For the transfer condition, the picture size displayed to the participants was derived as outcome variable.

## 6.12. EEG pre-processing

### 6.12.1. EEG device and feedback specifications

The neurofeedback device (NeuroConn THERA-PRAX) included a trainer screen and a client screen, and the neurofeedback apparatus (Item 4e of the CRED-nf checklist). Participants in the EEG group received the NF treatment using visual stimuli as reinforcement (Item 4c of the CRED-nf checklist). Specifically, two bar graphs depicting real-time beta band activity and muscular activity were presented (Item 4b of the CRED-nf checklist).

### 6.12.2. EEG pre-post recording

Changes in EEG spectral power during 180s resting-state eyes-open, 180s eyes-closed, and during 90s of food picture presentation will be analysed based on EEG recordings before the first and before the last neurofeedback session. A band-pass filter and a notch filter will be used to filter EEG data. The continuous EEG will be segmented in 2 s intervals and ocular artefact correction will be applied. Further, automatic artefact rejection will be applied to segments with voltage steps greater than 50 µV/ms and amplitudes exceeding ±100 µV. After these automated procedures, all data will be visually assessed by an extensively trained researcher in order to identify and remove residual contaminants (e.g., artefacts, focal abnormalities, drowsiness). Based on this assessment, at least 30 artefact-free segments (not necessarily consecutive) of the filtered EEG data will be required for the eyes-open and eyes-closed condition to be included in the analysis. The number of removed segments will be reported (Item 3e of the CRED-nf checklist). The filtered EEG data will then be Fourier transformed with a Hanning window length of 20% extracting theta (θ, 3.5–7.5 Hz), alpha (α, 7.5–12.5 Hz), and beta (β, 12.5–30 Hz) frequency bands. The extracted absolute power for each frequency band will be converted to

relative band power (%) and ln-transformed before statistical hypotheses testing to obtain normally distributed data.

### 6.12.3. EEG session-wise recording

During each session, beta band activity at baseline and during each trial were output from the neurofeedback device (in µV) and manually noted for the neurofeedback and transfer condition. For each participant, the number of successful trials, defined as trials with a beta band activity below the baseline activity, with a possible range between 1-12 (for neurofeedback) and 1-6 (for transfer) trials, and condition-wise (i.e., for the neurofeedback and transfer condition) contrasts between regulation and baseline will be provided.

## Publication bibliography

Carver, Charles S.; White, Teri L. (1994): Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. In *Journal of Personality and Social Psychology* 67 (2), pp. 319–333. DOI: 10.1037/0022-3514.67.2.319.

Fairburn, Christopher G.; Beglin, Sarah J. (1994): Assessment of eating disorders: Interview or self-report questionnaire? In *International Journal of Eating Disorders* 16 (4), pp. 363–370. DOI: 10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-#.

Garnefski, N.; Kraaij, V.; Spinhoven, P. (2001): Negative life events, cognitive emotion regulation and emotional problems. In *Personality and Individual Differences* 30 (8), pp. 1311–1327. DOI: 10.1016/S0191-8869(00)00113-6.

Garnefski, N.; Kraaij, V. Spinhoven, P. (2002): CERQ: Manual for the use of the Cognitive Emotion Regulation Questionnaire. Leiderdorp: Datec.

Görgen, Stefanie M.; Loch, Noelle; Hiller, Wolfgang; Witthöft, Michael (2015): Kognitive Emotionsregulation und Psychopathologie: Erste teststatistische Untersuchung des deutschen Cognitive Emotion Regulation Questionnaire (CERQ) in einer klinischen Stichprobe. In *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* 63 (4), pp. 255–265. DOI: 10.1024/1661-4747/a000248.

Gratz, Kim L.; Roemer, Lizabeth (2004): Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. In *Journal of Psychopathology and Behavioral Assessment* 26 (1), pp. 41–54. DOI: 10.1023/B:JOBA.0000007455.08539.94.

Hilbert, Anja; Tuschen-Caffier, Brunna (2016): Eating Disorder Examination – Questionnaire. 2. Auflage. Tübingen: dgvt-Verlag.

Kroenke, Kurt; Spitzer, Robert L.; Williams, Janet B. W.; Löwe, Bernd (2010): The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. In *General Hospital Psychiatry* 32 (4), pp. 345–359. DOI: 10.1016/j.genhosppsych.2010.03.006.

Meule, Adrian (2018): Food cravings in food addiction: exploring a potential cut-off value of the Food Cravings Questionnaire-Trait-reduced. In *Eating and weight disorders : EWD* 23 (1), pp. 39–43. DOI: 10.1007/s40519-017-0452-3.

Meule, Adrian; Hermann, Tina; Kübler, Andrea (2014): A short version of the Food Cravings Questionnaire-Trait: the FCQ-T-reduced. In *Frontiers in psychology* 5, p. 190. DOI: 10.3389/fpsyg.2014.00190.

Müller, A.; Smits, D.; Claes, L.; Zwaan, M. de (2013): Faktorenstruktur der deutschsprachigen version der BIS/BAS-Skalen in einer bevölkerungsstichprobe. In *Fortschritte der Neurologie-Psychiatrie* 81 (2), pp. 75–80. DOI: 10.1055/s-0032-1330482.

Santosa, Hendrik; Zhai, Xuetong; Fishburn, Frank; Huppert, Theodore (2018): The NIRS Brain AnalyzIR Toolbox. In *Algorithms* 11 (5), p. 73. DOI: 10.3390/a11050073.

Schwarzer, Ralf (Ed.) (1999): Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: R. Schwarzer.

Spitzer, R. L.; Kroenke, K.; Williams, J. B. (1999): Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. In *JAMA* 282 (18), pp. 1737–1744. DOI: 10.1001/jama.282.18.1737.

Spitzer, Robert L.; Kroenke, Kurt; Williams, Janet B. W.; Löwe, Bernd (2006): A brief measure for assessing generalized anxiety disorder: the GAD-7. In *Archives of internal medicine* 166 (10), pp. 1092–1097. DOI: 10.1001/archinte.166.10.1092.

Stieglitz, R.-D. (1999): Bullinger, M. & Kirchberger, I. (1998). SF-36. Fragebogen zum Gesundheitszustand. Göttingen: Hogrefe. Preis DM 298,-. In *Zeitschrift für Klinische Psychologie und Psychotherapie* 28 (2), pp. 143–145. DOI: 10.1026//0084-5345.28.2.143.

Strobel, Alexander; Beauducel, André; Debener, Stefan; Brocke, Burkhard (2001): Eine deutschsprachige Version des BIS/BAS-Fragebogens von Carver und White. In *Zeitschrift für Differentielle und Diagnostische Psychologie* 22 (3), pp. 216–227. DOI: 10.1024//0170-1789.22.3.216.

Teymoori, Ali; Gorbunova, Anastasia; Haghish, Fardzadeh E.; Real, Ruben; Zeldovich, Marina; Wu, Yi-Jhen et al. (2020): Factorial Structure and Validity of Depression (PHQ-9) and Anxiety (GAD-7) Scales after Traumatic Brain Injury. In *Journal of clinical medicine* 9 (3), p. 873. DOI: 10.3390/jcm9030873.

van Buuren, Stef; Groothuis-Oudshoorn, Karin (2011): mice : Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3). DOI: 10.18637/JSS.V045.I03.