Supplementary Online Material

Alcohol-Specific Inhibition Training in Patients with Alcohol Use Disorder: A Multicenter, Double-Blind, Randomized Clinical Trial Examining Drinking Outcome and Working Mechanisms.

Maria Stein, Ph.D.; Leila M. Soravia, Ph.D.; Raphaela M. Tschuemperlin, Ph.D.; Hallie M. Batschelet, Ph.D.; Joshua Jaeger, M.Sc.; Susanne Roesner, Ph.D.; Anne Keller, M.D.; Juan Martin Gomez Penedo, Ph.D.; Reinout Wiers, Ph.D.; & Franz Moggi, Ph.D.

2022

This supplementary material has been provided by the authors to provide readers with additional information about their work.
Note that there is additional information (including all code used in the analyses of the primary outcome) available on open science framwork:
https://osf.io/6akfd/?view_only=d68b05eb4b414b619ac56a4d4fe8de4b

1 Methods: additional details

1.1 Outcome measures and study arms as listed in the trial registration

1.1.1 Outcome measures

This trial's registration at clinicaltrials.gov (NCT02968537) lists ten outcomes, five of them as primary, and five as secondary outcomes. At the time of registration, we were not aware of the CONSORT guidelines requiring a focus on one single primary outcome (1). In order to follow the CONSORT guidelines, we thus slightly deviate from our registry here; We now consider only our most important primary outcome, the percentage of days abstinent (PDA, which was listed first in the trial registration as well as in the study protocol), as our primary outcome measure. All other outcomes are now considered secondary outcomes.

Not all ten outcomes listed in the registration are reported in the present manuscript. The following list informs the reader on the reporting of each of the outcomes:

Primary outcome:

Percentage of days abstinent (PDA) at 3-months follow-up: This outcome is reported here, see section outcome measures of the main article and section 1.5.2 of the SOM for further details.
Note that as a primary outcome variable, 3-months follow-up data was preferred over longer follow-up intervals because this is the first clinical trial on Alc-IT, and non-clinical studies assessed even short follow-up intervals. Thus, the stability of Alc-ITs effects remains unknown. In this situation, a 3-month follow-up assessment provides clinically relevant data with minimal risk to miss effects because they were assessed too late.

Secondary outcomes:

Inhibitory control as assessed by the change in errors of commission in the Go-NoGo task from pre- to post-training assessment: This outcome is reported here, see the main article and sections 1.4.1 and 2.5 of the SOM for further details.

Neurophysiology of inhibitory control assessed by the N2/P3-components of the event-related potentials (ERPs) computed from electroencephalogram (EEG) measurement during Go-NoGo task at pre- and post-training assessment: This outcome is not reported here, but will be reported in a separate article, because it exceeds the scope of the present article. ERPs were assessed in a subgroup of patients included in the present study, as well as in an additional healthy control group (2), warranting a separate description of the sample along with a substantial amount of methodological information required for thorough reporting of an ERP study.

Implicit associations as assessed by the change in d-score calculated from the implicit associations test (IAT) from task from pre- to post-training assessment: This outcome is reported here, see the main article and section 1.4.2 of the SOM for further details.

Percentage of heavy drinking days (PHDD) at 3 months follow-up: This outcome is reported here. Note however that instead of the number of heavy drinking days (as listed in the registration), we analyzed the percentage of heavy drinking days (PHDD). This allowed us to control for days spend in a protected environment. See section outcome measures of the main article and section 1.5.2 of the SOM for further details.

Abstinence-related self-efficacy as assessed with the alcohol abstinence self-efficacy scale (AASE, 3): This outcome is not reported, because observations of study team members during data collection as well as inspection of raw data suggested that a relevant subgroup of study participants might have had trouble understanding the questions in this self-rating scale. Thus, the validity of this data is seriously hampered, and this outcome was excluded from the study.

Craving as assessed by change in the obsessive-compulsive drinking scale (OCDS-G, 4) from baseline to 3-month follow-up: This outcome is reported here, see section 2.6 of the SOM for further details. OCDS-data from 6- and 12-month follow-ups will be reported elsewhere.

Time to first drink (TTFD) at 3-month follow-up: This outcome is reported here, see the main article and sections 1.5.2 and 2.4 of the SOM for further details. TTFD-data from 6- and 12-month follow-ups will be reported elsewhere.

Percentage of heavy drinking days as well as the percentage of days abstinent at 6- and 12-month follow-ups: Will be reported elsewhere.

Inhibition, as assessed by the change in stop-signal response time (SSRT), calculated from the stop-signal task (SST) during pre- and post-training assessment: To increase comparability with important prior studies (5), we used an SST with fixed latencies. Data inspection of SST data suggested that this task design led to the SST being too hard for our sample and that the premises necessary to calculate the SSRT were not met in some of our patients. This problem is explained in 1.4.3 and hampered inferences about training effects on SSRT.

1.1.2 Study arms: training condition and daytime of training

Next to the 3 training groups (standard Alc-IT, improved Alc-IT, and control), the present study also investigated whether the daytime of training influenced training effects. This was motivated by the rationale that endogenous cortisol, a consolidation enhancer, is higher in the morning than in the afternoon and might enhance training effects, which rely on learning processes. Therefore, part of our sample completed their allocated training version in the morning (within 2 hours after awakening) while another part participated in their allocated training version in the afternoon. The daytime of training was also randomly assigned. The results regarding this secondary research question are reported in the main manuscript.

1.2 Additional information concerning the stimulus material

The stimulus material for the three training versions, the Go-NoGo task (GNG) and the implicit association test (IAT), were identical and consisted of eight pictures of alcoholic beverages (either eight beer, eight wine, or eight spirits were used), eight pictures of water and eight neutral objects (e.g., stapler). All pictures were taken against a white background and controlled for size (344 × 400 pixels) and luminance (6).

1.3 Additional information on the training interventions

In all three training versions, participants were instructed to answer as fast and as accurately as possible: They were asked to press a button when a Go-cue (the letter p or f, counterbalanced) appeared in one of the four corners of the picture and to withhold from responding when a NoGo-cue (“f” or “p”) appeared instead. Picture and cue appeared simultaneously and were displayed for 1500ms.
In both versions of the alcohol-specific inhibition training (Alc-IT,) pictures of alcoholic beverages were consistently paired with a NoGo cue, while Go cues were distributed among the other picture types

(water, neutral). In contrast, in the control training, a nonspecific inhibition training, all three picture types appeared equally often with NoGo- and Go-cues:

Standard Alc-IT used a Go/NoGo ratio of 50/50 (as developed by Houben et al. (7)). Originally, this standard Alc-IT included 80 alcohol-related NoGo trials and 80 water Go trials. In order to match both Alc-IT versions in training length while keeping the number of alcohol-NoGo pairings constant, an additional 80 neutral Go trials and 80 neutral NoGo trials were included.

Improved Alc-IT operated with a Go/NoGo ratio of 75/25. It equally included 80 alcohol-related NoGo trials and 80 water Go trials, but contrary to standard Alc-IT, 160 neutral Go trials completed this training. The development of improved Alc-IT is based on the rationale that a task with a higher Go/NoGo ratio targets the inhibitory system in a more reliable way (8) or at least increases the inhibitory demands during the task performance. Furthermore, effect sizes describing inhibitory deficits in AUD were higher for tasks with higher Go/NoGo ratios as compared to equiprobable tasks (9). Thus, improved Alc-IT was designed to pose higher demands on the inhibitory system, aiming to enhance inhibitory training effects.

Both Alc-IT versions thus were of equal length and comprise equal numbers of Alcohol-NoGo pairings,but differ in the Go/NoGo ratio and thus in inhibitory control difficulty.

Control training also comprises 80 alcohol-related, 80 water, and 160 neutral trials, but all picture types appeared equally often in Go- and in NoGo-trials. Thus, this control training is a nonspecific inhibition training, that matches the two alcohol-specific training interventions in terms of length and picture sets used.

All three versions were presented with Inquisit. The original inquisit script from Houben et al. (7) was adapted to incorporate not only the standard Alc-IT, but also the improved Alc-IT and control training. A preprogrammed sequence further guaranteed that participants were directed to the correct training version (as assigned by the randomization list) when entering their subject number.

1.4 Additional information concerning the experimental tasks

Administration and response logging of all tasks was performed with E-Prime 2.0 (EP2Pro2.0.10.356, Psychology Software Tools, Sharpsburg, PA, USA).

1.4.1 Additional details on the Go-NoGo Task

Task description: The Go-NoGo-task (GNG) assesses the action restraint component of response inhibition. It is thus conceptually close to the training versions, which are also GNG-based but differ in precise instructions, timing, and NoGo-cues. The stimulus material was identical to the alcohol-related and water stimuli used during training and consisted of eight pictures of alcoholic beverages (either beer, wine, or spirits; tailored to the participants' drink of choice) and eight pictures of water.

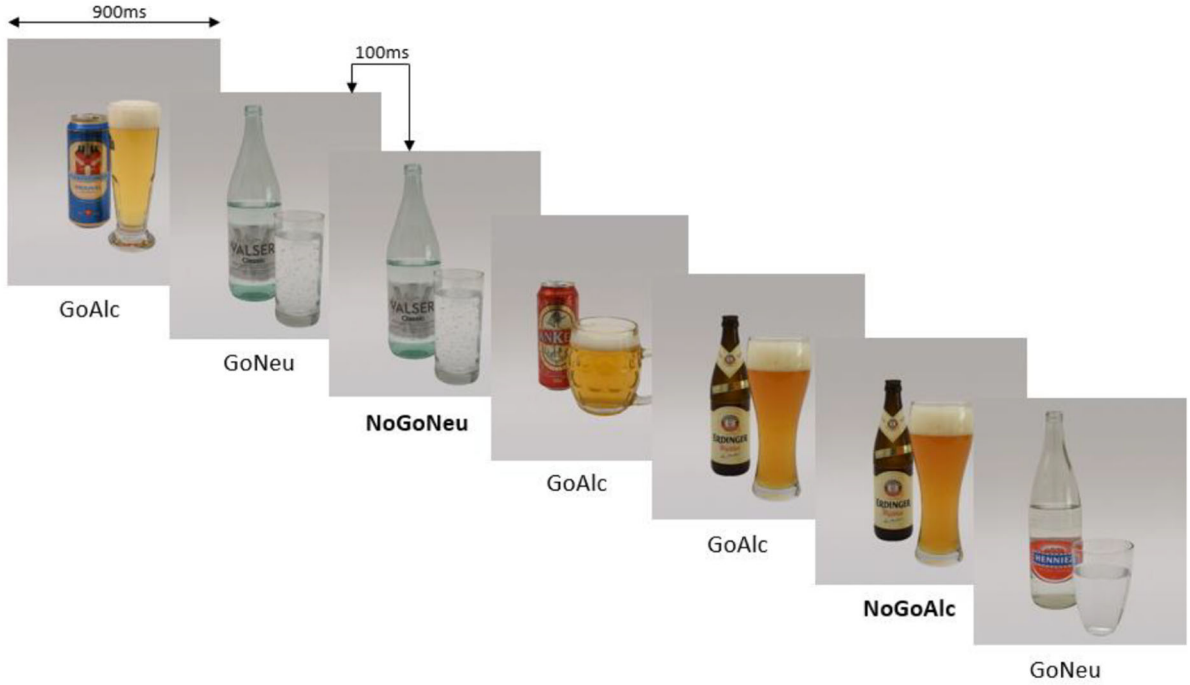
Pictures of alcoholic beverages or water were presented on the screen for 900 ms with a 100 ms interstimulus interval (eFigure 1). Participants had to press a button as fast as possible whenever a stimulus appeared on screen (Go trials), unless a stimulus was repeated (NoGo trials). Each of the eight alcohol-related and eight water pictures served 52 times as a Go trial and eight times as a NoGo trial. The resulting Go/NoGo ratio of 6.5 established a prepotent response tendency, which, in combination with the relatively fast pace of the task, made inhibition effortful. The 960 trials (416 neutral Go trials, 416 alcohol-related Go trials, 64 neutral NoGo trials, and 64 alcohol-related NoGo trials) were presented in a pseudo-randomized order, controlling for sequential order, position effects, and guaranteeing a minimum of two Go trials between two NoGo trials. In the middle of the task, participants were allowed

a pause of a self-determined length. Before starting the GNG task, a training version using four pictures of household items (e.g., stapler) was used to familiarize participants with the procedure.

Data Analysis: Errors of commission (EOCs, i.e., incorrect button presses) in NoGo trials and errors of omission in Go trials were counted separately for alcohol-related and neutral trials in each individual and were entered as the dependent variable in the non-parametric analyses investigating a potential inhibitory working mechanism. These analyses were performed in R using the nparLD package (10), which provides non-parametric ANOVA-type statistics (ATS) for multifactorial designs. The first overall analysis used the f1.ld.f2 function and evaluated the effects of the within-factors time (pre, post) and picture type (alcohol-related, neutral) and the between-factor training (standard Alc-IT, improved Alc-IT, Control) as well as their interactions. Following significant interactions, follow-up analyses used the f2.ld- or f1.ld-functions, respectively, to assess the effects of time and picture type separately in each training group.

Reaction times (RTs) were averaged in each individual for alcohol-specific and neutral Go-trials separately after excluding RTs below 150ms, because those are highly unlikely to represent voluntary responses on such a visual task (11) and might as well represent delayed responses to the previous stimulus in a fast-paced task. Mean RTs are presented for descriptive purposes (see, eTable 5; SOM section 2.5).

eFigure 1: Go-NoGo-task with alcohol-related and neutral Go- and NoGo-trials



Note. Pictures were presented on screen for 900ms each, with a 100ms inter-stimulus interval. Participants were instructed to press a button whenever a new stimulus appeared on screen (Go-trial), unless the same stimulus was repeated twice (NoGo-trial). Abbreviations: GoAlc: Alcohol-related Go trial; GoNeu: Neutral Go trial; NoGoAlc: Alcohol-related NoGo trial, NoGoNeu: neutral NoGo trial.

1.4.2 Additional details on the Implicit Association Task

Task description: The Implicit Association Task (IAT) assessed the automatically activated implicit associations to alcohol and was used to monitor the potential effects of Alc-IT on the valence automatically assigned to alcohol-related stimuli. It thus monitored the effects, which would be expected following the stimulus devaluation hypothesis.

The stimulus material for the IAT consisted of pictures of either alcohol or water and positive and negative adjectives. The eight pictures of alcoholic beverages (beer, wine, and spirits, depending on the participant's drink of choice) and eight pictures of water were identical to those used during training. The affective adjectives consisted of eight positive (happy, jolly, energetic, funny, sociable, attractive, cheerful, smart) and eight negative (dull, miserable, sick, depressed, unhappy, disgusting, angry, foolish) adjectives. Affective adjectives were derived from Houben et al. (7), translated into German and linguistically validated (6).

In this alcohol valence IAT, during each trial, one of the stimuli (either a picture or a word) appeared on the screen in a pseudorandomized order. Participants were instructed to indicate via button press, as fast and as correct as possible, to which of the target categories (alcoholic beverages vs. water) or affective categories (positive vs. negative adjective) the stimulus belonged. Buttons "a" on the left side of the keyboard and "l" on the right side of the keyboard served as response buttons, assignments were counterbalanced. Stimuli appeared in the center of the screen for a maximum of 1750 ms or until a response was recorded, followed by feedback (correct or false) for 200 ms and an interstimulus interval of 250 ms.

During "alcohol-positive" blocks, stimuli from the alcohol and positive categories shared one response button (e.g. "a" on the left side of the keyboard), while water and negative shared the other response button (e.g. "l" on the right side of the keyboard). Contrarily, during "alcohol-negative" blocks, stimuli from the alcohol and negative categories were assigned to the same response button, while water and positive attributes shared the other one. As a reminder of the current assignment instructions, category labels were displayed in the upper two corners of the screen throughout the IAT (eFigure 2).

Two versions of the IAT, each consisting of 14 blocks (12, 13), were created so that starting blocks, as well as key assignments, could be counterbalanced and sequence effects controlled for (see Tschuemperlin et al., 2020 (13), for an overview of IAT block order). At the beginning and with every change in key assignment, practice blocks (16 trials) allowed participants to familiarize themselves with the key assignments: First, the classification of the target concepts (alcohol, water) and/or the affective categories (positive, negative) to left ("a") or right ("l") button was practiced. Then, participants practiced the combined assignment of target concepts and affective categories to one button (e.g., in alcohol-negative blocks: press "a" when an alcohol-related picture or a positive word appears and press "l" when a water picture or a negative word appears). After these practice blocks, a test block of 64 trials followed. The IAT incorporated four test blocks, two with alcohol pictures and positive words sharing a response button (alcohol-positive blocks) and two with alcohol pictures and negative words sharing a response button (alcohol-negative blocks). Rest periods of self-determined length were allowed between all blocks.

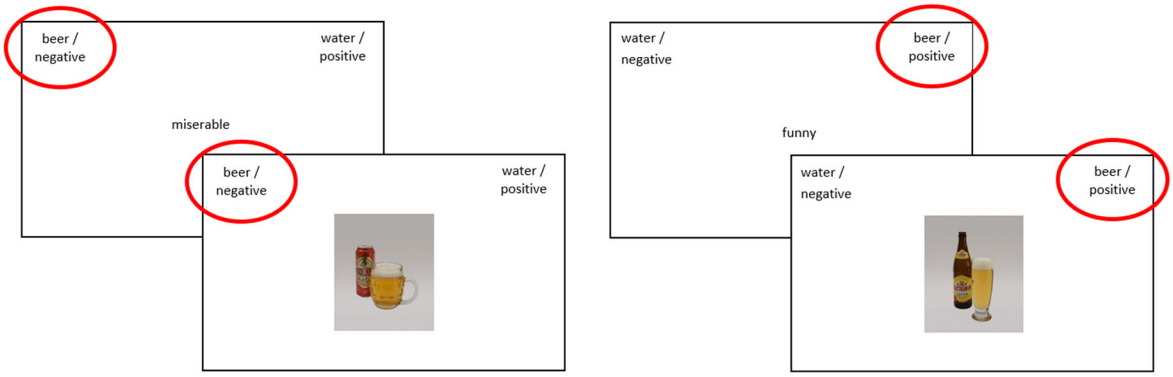
Data Analysis (calculation of d-score from the IAT-data): Building on the idea that reactions are easier and thus faster, if associated concepts (e.g., alcohol and positive attributes) share one response button, than if less associated concepts (e.g., alcohol and negative attributes) have to be assigned to the same response button, reaction time (RT) differences between alcohol-positive and alcohol-negative blocks served as an indicator of implicit alcohol associations.

Similar to earlier studies (12, 13) behavioral analyses concentrated on the 160 alcohol trials (alcohol-positive assignments and alcohol-negative assignments) and RT differences were calculated between these trials from alcohol-positive and alcohol-negative blocks. These RT differences were calculated as d-scores as described by the improved scoring algorithm (14) and using the same R-Script as in Tschuemperlin et al. (2020) (13), which implemented the following steps:

1) All trials with RT above 300 ms from the combined blocks were used. Participants with more than 10% of RTs under 300 ms (n = 9) were excluded. Incorrect trials were replaced with individuals' mean RT of correct trials with 600 ms added.
2) Individual means of practice and test blocks, as well as a pooled standard deviation (SD) for practice trials and trials from test blocks, were calculated.
3) Alcohol-positive trials were subtracted from alcohol-negative trials for practice and test blocks separately, normalizing these differences by the individualized pooled SD of practice or test block, respectively.
4) Finally, the two quotients were averaged to compute the alcohol-related IAT effect score (dALC), which quantifies the individual alcohol valence bias: the higher a participant's d-scores, the more positive implicit associations to alcohol this participant holds.

Complete IAT data (from pre- and post-training sessions) were collected from 223 subjects. 9 subjects had to be excluded from IAT-analyses (see point 1 above), leaving a final analytic sample of 214 subjects (N=72 in standard Alc-IT, N=73 in improved Alc-IT, N=69 in Control Group) for the IAT analyses.

eFigure 2: Implicit Association Task (IAT)



Note. Overview of exemplary trials in the alcohol-positive and alcohol-negative blocks of the IAT, beer version. For the wine and spirits version, the respective stimulus set was used.

1.4.3 Details on the Stop signal task (SST)

Stimulus material and task description: The SST was included in the study to assess the action cancellation component of response inhibition. This was done to investigate whether potential effects of the Go-NoGo-based training generalize to a different, albeit related task. Stimuli material used in the SST differed from the stimuli used during training and during IAT and GNG: For the SST, we used pictures of alcoholic beverages (either beer, wine, or spirits) and neutral objects taken in everyday life, which were chosen from the stimulus pool established and controlled by Fey et al. (15). Four pictures of neutral objects located outdoors (e.g. wheelbarrow) and four pictures of alcoholic beverages taken indoors (either 4 beer, 4 wine, or 4 spirits pictures, tailored according to the drink of choice) served as stimuli. As these stimuli differed from the ones used in the training, an effect in the SST would also

indicate generalization to new stimuli. Due to limited time and resources, these two aspects of generalization could not be investigated separately.

During the SST, whenever a picture appeared on screen, participants were instructed to indicate as fast as possible by pressing a letter on the keyboard (letters “a” and “l”, counterbalanced) whether the picture was taken indoors or outdoors. Instructions did not mention the deliberate confounding of picture location (indoor/outdoor) with content (alcohol-related/neutral).

Each trial started with a fixation cross displayed for 460ms, after which one of the pictures appeared on screen, prompting participants to initiate the correct button press. The picture remained on screen for 1500ms or until a response was given. In 25% of the trials (the NoGo- trials), an auditory stop signal occurred, starting with a delay of either 100, 200, or 300ms (each delay used in one third of the Stop-trials), and prompting participants to cancel their response. Fixating the delay and using the latencies of 100ms, 200ms, and 300 ms was done to increase comparability with non-clinical studies (5), however, a delay of 400ms was deemed too hard for our clinical population and omitted. After two practice blocks, to familiarize with the task, the two test blocks, each composed of 240 trials presented in pseudorandomized order (180 Go and 60 NoGo trials, half of each alcohol-related, half neutral), were administered.

We attempted to compute the Stop-signal reaction time (SSRT) as described in the consensus guidelines (16) for SSTs without tailoring of the delay of the cancelation signal. These guidelines also state that the SSRT should not be computed, when the conditional probability p(respond|stop-signal) significantly deviates from 0.5, specifically, it states that SSRT should not be computed with a conditional probability is below 0.25 or higher than .75. Unfortunately, for the delay of 300ms, this conditional probability was p(respond|stop-signal) ≥0.9 for a majority of our subjects (all but 4 of the subjects). Unfortunately, we could therefore not compute a reliable SSRT.

Regarding the SSRT, the only conclusion to be taken from this study is thus that difficulty has to be lowered for patient populations and that usage of the tracking methods, which allows to tailor the difficulty in response to the error rate of a given individual (16), is highly recommendable for future studies.

1.5 Additional information on questionnaires and interviews

1.5.1 Measures used for descriptive purposes and confirmation of eligibility

At baseline, relevant demographics (e.g., age, gender, education), information on previous AUD treatment(s), and other mental health problems were assessed. The AUD diagnosis was verified using the German version of the semi-structured interview of the Diagnostic Expert System for Psychiatric Disorders (DIA-X (17), adapted to the Diagnostic and Statistical Manual of Mental Disorders [DSM-5] specifically for AUD). Additionally, the Alcohol Use Disorder – Scale (AUD-S (18), adapted to DSM-5), assessed self-rated AUD symptoms, and the Alcohol Use Disorder Identification Test (AUDIT (19)), evaluated the severity of problematic drinking. Other substance abuse (Drug Use Disorder Identification Test, DUDIT (20)), motivation to change drinking (taking-steps subscale of SOCRATES (21)), craving (Obsessive-Compulsive Drinking Scale, OCDS (4)), and psychopathological symptoms (general symptom index of the Brief Symptom checklist, BSCL-GSI (22)) were measured and are reported in the main paper.

To provide a more comprehensive description of our sample, traits of antisocial personality disorder (ASPD, assessed with an questionnaire adapted from (23)), symptoms of attention-deficit and hyperactivity disorder (ADHD, assessed using the Adult ADHD Self-Report Scale [ASRS (24)], anxiety

(Beck Anxiety Inventory [BAI] (25)), depressive symptoms (Beck's Depression Inventory [BDI] (26)), behavioral inhibition and avoidance (Behavioral Inhibition System/Behavioral Approach System Scale [BISBAS] (27)), impulsivity, (Scale for Impulsive Behavior – I8 (28)), sensation seeking (Need Inventory of Sensation Seeking [NISS] (29)), as well as stress and coping (stress and coping inventory [SCI] (30)) were measured. In addition, motivation and confidence to achieve treatment goals were additionally assessed with single questions (31, 32) to which patients answered on an 11-point Likert scale. Moreover, treatment goals were assessed as multiple-choice questions (32). All these additional measures are provided in the SOM in eTable 1. See Tschuemperlin et al.(6) for a complete list of measures obtained at each time point.

1.5.2 Measures of alcohol consumption (including the primary outcome)

Alcohol consumption was assessed threefold. (I) The TLFB, which is the gold-standard for assessment of alcohol use, provided the primary outcome. Additional indicators of alcohol use were collected with the (II) HDL-questionnaire and (III) a short interview to provide a broader assessment of alcohol use. These two measures were only included in the secondary analysis consisting of hierarchical linear models (HLMs, see section 2.3). Details on all three measures follow:

I. *The Timeline Followback* (TLFB) (33) was administered in an interview (face-to-face at baseline and by telephone at 3-month follow-up) by a trained member of the study team. The TLFB at baseline assessed drinking during the 90 days prior to detoxification entry, whereas the TLFB at 3-month follow-up assessed drinking during the 90 days following discharge from treatment. Alcohol consumption was assessed in standard drinks (StD) consumed each day, with one StD equating to 3 dl beer, 1 dl wine, 2 cl spirits/liquor, or 10 g of pure alcohol. The percentage of days abstinent (PDA) and percentage of heavy drinking days (PHDD) were computed for the time period assessed. PDA computed from TLFB-data assessed at 3-month follow-up provided the primary outcome measure for this study.

Data reduction for primary outcome measure PDA: PDA was calculated as the percentage of days without alcohol consumption (PDA = (days abstinent / 90) × 100). Because staying in a protected environment (defined as psychiatric, substance-related, or somatic inpatient treatment) can strongly distort this measure, the formula was adjusted for patients with such an inpatient stay. The adjustment had the effect to exclude these days, during which alcohol use was not possible or highly unlikely (PDA = ((days abstinent – days inpatient treatment/(90 – days in inpatient treatment) × 100).
Of the total sample of 242, 35 (14%) at baseline and 18 (7%) at 3-month follow-up reported an inpatient stay and were submitted to the adjusted formula. Note that the proportion of patients with such an inpatient stay did not differ significantly between the three training groups.

Data reduction for secondary outcome PHDD: PHDD was also computed for alcohol consumption during the last 90 days as the percentage of days on which alcohol consumption exceeded 4 (women) or 5 (men) StD, respectively (PHDD = (heavy drinking days/90)×100). Again, to prevent data distortion due to staying in a protected environment (defined as psychiatric, substance-related, or somatic inpatient treatment), the formula was adjusted for those patients with such an inpatient stay (PHDD = ((heavy drinking days / (90 – days in inpatient treatment) × 100).
Same as for the analyses of PDA, 35 (14%) of the total sample at baseline and 18 (7%) at 3-month follow-up reported an inpatient stay and were submitted to the adjusted formula, with no differences in the proportion of patients with an inpatient stay between the three training groups.

Data reduction for secondary outcome time to first drink: Time to first drink was drawn from the TLFB data and was defined as the first day after treatment discharge on which alcohol was consumed. Only patients who were not living in a protected environment following treatment discharge and who provided an indicator of time to first drink in the TLFB (n=156) were retained in these analyses.

II. *The Health and Daily Living Form* (HDL) (34) was used as a questionnaire to assess patients' alcohol use in the last 90 days prior to their current detoxification (at baseline) or in the 90 days following discharge from treatment (at 3-month follow-up). In HDL, patients report consumption frequency and quantity of beer, wine, liquor, and spirits, resulting in a total number of StD consumed per day. Separate items assess the PDA and PHDD in the last 90 days. HDL data was only used in the additional HLM analyses reported in SOM, (section 1.7.2 and 2.3). In order to control for distortion due to an inpatient stay, information about such a stay was drawn from the TLFB

III. *A short interview* (SI) at the 3-month follow-up was conducted on the telephone as soon as the first contact could be established with the patient. During this short interview, patients were also asked to indicate their PDA and PHDD. Data from the short interview was only used in the additional HLM analyses reported in SOM (section 1.7.2 and 2.3).

1.6 Additional information about the statistical analyses of group differences

1.6.1 Assessment of potential differences between training groups

One-way ANOVA was used to assess potential differences in age, prior detoxifications, education, AUD-S, OCDS, BSCL-GSI, and SOCRATES-TS between the training groups at baseline. Potential group differences in civil status, gender, and treatment goals were assessed using a chi-square test.

1.7 Additional information about statistical assessment of training effects

1.7.1 Main analysis of training effects on primary outcome PDA and secondary outcome PHDD: Treatment of missing data

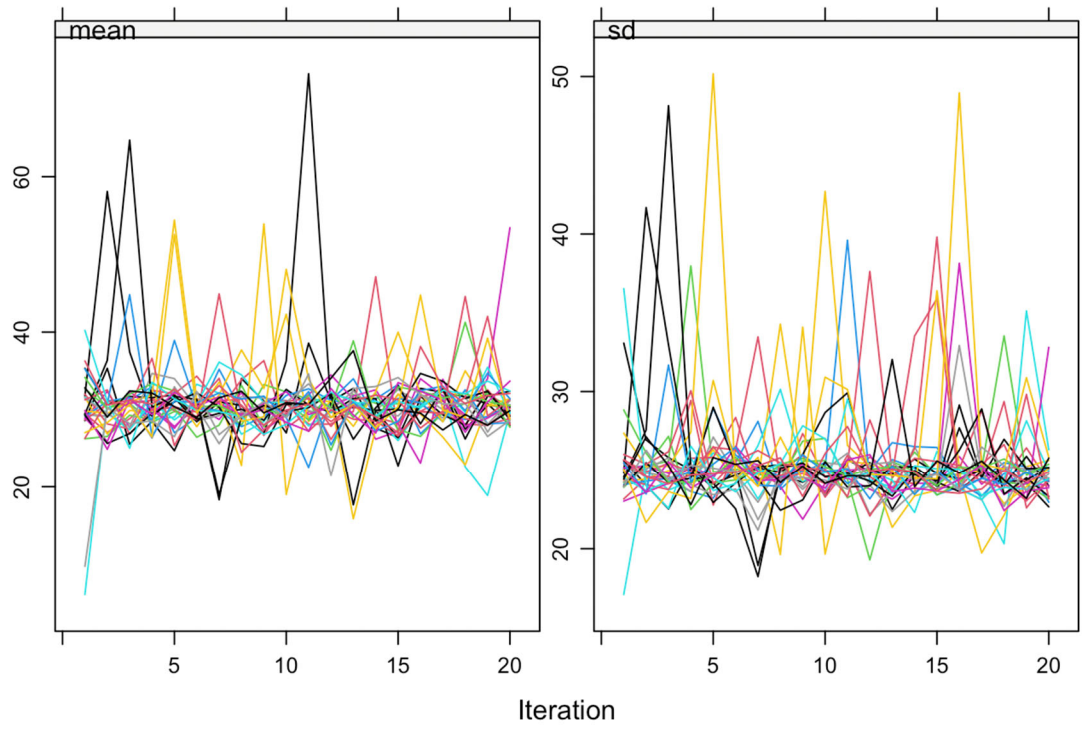
Treatment of missing data. To analyse whether TLFB outcome values are missing not at random (MNAR), at random (MAR), or completely at random (MCAR)—we compared treatment groups regarding the number of missing data, conducted Little's MCAR-Test, and evaluated if there is a relationship between the propensity of a TLFB value to be missing and its values.
At the 3-month follow-up, n=69 (28.5%) patients of the total sample (n=242) did not provide TLFB data. No significant differences between the three training intervention groups regarding the number of missing observations were observed ($p > .75$). Little's MCAR-Test was significant ($\chi 2$ (69, N = 242) = 41.00, $p = .0012$), pointing in the direction of not MCAR. Thus, we did not adopt an MCAR assumption as potential associations between missing outcome values and specific patient variables are plausible:
To distinguish between MNAR and MAR, and therefore evaluate if the probability of missingness depends on the unobserved missing values themselves, we compared additionally collected data of TLFB non-respondents to that of TLFB-respondents. To do so we used PDA estimates of TLFB non-respondents obtained through a short telephone interview (SI; n = 11) or a questionnaire (HDL; n =16). In a regression analysis with a combined sample of n = 200 and TLFB-respondent/non-respondent as a dichotomous predictor, we compared the SI- and HDL-PDA values of the non-respondent with the

TFLB-PDA values of the respondents and found no evidence of a difference in their outcome probability distributions (β = -3.22, $t_{(199)}$ = -0.59, p = .56).

Given these results, it seems reasonable to us to assume that the TLFB outcomes are MAR. Following the guidelines of Jakobsen et al. (2017) (35) we performed multiple imputations by chained equations using the "MICE" package in R. In line with White et al. (2011) (36), we choose the number of imputations to be 30 (i.e. higher than the percentage of cases that are incomplete). The incomplete outcomes were imputed under a fully conditional specification. The dataset contained: a) the outcome variable measured with the TLFB (i.e., PDA or PHDD at three-month follow-up), b) predictor variables (i.e., training intervention; daytime of training), and c) covariates of interest (i.e., site; age; gender; pharmacotherapy; PAD/PHDD at the baseline). Parameters were estimated separately in each imputed data set and subsequently combined (pooled) using Rubin's rules. By calculating the means and standard deviations of the imputed values in each cycle, the stability of the imputed data and the presence of successful convergence were visually evaluated (see, eFigure 3).

In addition to analyzing the imputed data under the MAR assumption, we also performed sensitivity analyses under two alternative assumptions: a) MNAR, under which missing TLFB values of subjects lost to follow-up were replaced with that subject's previously observed value—i.e., Last Observation Carried Forward (LOCF); and b) MCAR, under which TLFB responders are assumed to be representative of the entire sample—i.e., Complete-case analysis (see, SOM 2.2). Even if these analyses provide important data about the stability of the effects, one has to acknowledge, that their results are probably more biased than the main analysis (37). As a further analysis, we also analysed complementary alternative measurements of the outcome variables in hierarchical linear models (see, SOM 1.7.2 and 2.3).

eFigure 3: Mean and standard deviation of imputed PDA values



1.7.2 Additional analysis of treatment effects on PDA and PHDD: hierarchical linear models (HLMs)

A series of models compared standard Alc-IT and improved Alc-IT against the control condition. See section 2.3.2 for the equations implemented in these models. All HLM analyses were run with R-Studio (Version 1.2.5033, RStudio Team, 2015), an integrated development environment for the R programming language (version 3.5, R Core Team, 2019). The lme4 R package (38) was used to compute the HLMs, and gglpot2 (39) was used to visualize the results. The reason for this additional analysis using a hierarchical structure was the possibility to integrate additional, alternative drinking outcome data measured with the Health and Daily Living Form (HDL) and a short interview (SI) in a combined model.

1.7.3 Analyses of treatment effects on drinking: Time to first drink

The Kaplan-Meier estimator was used to estimate a non-parametric statistical survival function to determine the relative probability that an individual patient will be abstinent past a given number of days (d). At d = 0, the Kaplan-Meier estimator is 1 and approaches 0 as d progresses. Patients who remained abstinent until the 3-month follow-up assessment were censored. A log-rank test derived from a Cox proportional hazards model (40) was used to compare the three training intervention groups

1.7.4 Analyses of treatment effects: Errors of commission on Go-NoGo-task

GNG data were obtained from 224 participants. Because of a non-normal distribution, data were analyzed with non-parametric analyses provided in the package nparLD in R, yielding ANOVA-type statistics (ATS) for each effect (10). Note that this presents a deviation from the original analysis plan (6) where we omitted to include a non-parametric analysis option for the case that our secondary outcomes were not normally distributed and thus did not mention this option to analyze our multi-factorial design non-parametrically with ATS. As GNG data were not normally distributed, ATS represents a more adequate statistical method for our data.

First, a three-factorial analysis assessed the effects of the factors training group (standard Alc-IT, improved Alc-IT, Control), time (pre, post), and picture type (alcohol-related, neutral) on errors of commission (EOCs) using the f1.ld.f2 function. A significant interaction was followed up with two-factorial analyses to assess the effects of time and picture type in each group separately using the ld-f2 function and finally with one-factorial analyses assessing the effect of time for each picture type separately using the ld-f1 function.

2 Results: additional and more detailed

2.1 Description and comparison of sample characteristics across treatment groups.

A selection of clinical and demographic characteristics was reported in the main paper. Note that for none of the baseline characteristics reported in the main paper, differences between treatment groups were detected (for age and days in residential treatment: all $F<0.07$, all $p>0.9$; for all other variables: all $\chi^2< 11.5$, all $p>0.32$).
Here, we provide additional parameters to describe the sample (eTable 1). Note that also for none of these additional characteristics, differences between treatment groups were detected. In addition, no difference in the amount of data available at baseline or at the 3-month follow-up measurement between the three training groups was detected.

Treatment programs in residential treatment were planned to last approximately 8 to 12 weeks. While 70% of our sample completed their inpatient stay within this range, 10% had a shorter stay and 20% stayed longer. The median value of days of residential treatment was 84 days, for mean, range and standard deviation see table 2 of the main manuscript.

eTable 1: Additional sample characteristics

Variable	Training Group												Statistics		
	Total sample			Control			Standard Alc-IT			Improved Alc-IT					
	(*N* = 242)			(*n* = 79)			(*n* = 84)			(*n* = 79)					
	n	M	SD	n	M	SD	n	M	SD	n	M	SD	df	*F*	*p*
Motivation at discharge															
To drink less	216	9.50	1.10	68	9.59	.85	73	9.53	.88	75	9.40	1.44	2	.560	.572
To get problem under control	219	9.44	1.25	69	9.55	.93	74	9.42	1.11	76	9.37	1.60	2	.401	.670
Motivation (3-month FU)															
To drink less	194	9.33	1.66	58	9.31	1.74	66	9.18	1.88	70	9.48	1.33	2	.548	.579
To get problem under control	191	9.20	1.58	58	9.19	1.42	66	9.06	1.78	67	9.36	1.49	2	.595	.553
Confidence at discharge															
Abstinent in 12 mo.	193	8.14	1.93	60	8.20	1.78	65	8.35	1.92	68	7.90	2.07	2	.934	.395
Controlled drinking in 12 mo.	37	8.32	2.04	10	8.40	1.78	15	8.00	2.53	12	8.67	1.61	2	.352	.706
Confidence (3month FU)															
Abstinent in 12 mo.	157	8.09	2.01	45	7.84	2.38	56	8.39	1.66	56	7.98	2.00	2	1.055	.351
Controlled drinking in 12 mo.	25	8.40	1.38	10	7.90	1.60	7	8.71	.951	8	8.75	1.39	2	1.097	.351
BISBAS															
BIS	240	2.96	.54	79	2.93	.54	82	3.00	.54	79	2.95	.55	2	.394	.675
BAS	238	2.89	.40	79	2.91	.40	83	2.88	.43	76	2.88	.39	2	.114	.892
I8	237	2.87	.55	79	2.87	.56	82	2.91	.53	76	2.84	.56	2	.350	.705
NISS	242	2.69	.58	79	2.70	.62	84	2.70	.52	79	2.66	.61	2	.136	.873
ASRS															
Inattentive	242	15.93	5.73	79	15.52	5.70	84	16.46	5.93	79	15.78	5.57	2	.592	.554
Impulsive / Hyperactive	242	14.05	5.77	79	13.76	5.98	84	14.65	5.90	79	13.70	5.44	2	.708	.494
ASP	238	8.97	7.22	79	8.67	7.61	82	8.83	6.82	77	9.42	7.26	2	.228	.796
BAI	242	10.14	9.89	79	8,78	7.99	84	11.15	10.44	79	10.43	10.94	2	1.219	.297
BDI	241	14.72	10.30	79	14.08	9.24	83	15.73	11.90	79	14.30	9.50	2	.621	.538
SCI	238	2.54	.41	77	2.54	.38	82	2.51	.42	79	2.56	.42	2	.319	.727
	n	%		n	%		n	%		n	%		df	χ2	*p*
Therapy goal at discharge															
None	7	3.3		2	3		3	4.1		2	2.7		8	3.940	.862
Abstinence	89	41.8		27	40.3		34	46.6		28	38.4				
Limited Abstinence	97	45.5		31	46.3		29	39.7		37	50.7				
Controlled Drinking	20	9.4		7	10.4		7	9.6		6	8.2				
Therapy goal at 3-month FU													4	.817	.936
None	10	5.1		2	3.4		3	4.4		5	7.1				
Abstinence	123	62.8		35	60.3		43	63.2		45	64.3				
Limited Abstinence	35	17.9		11	19		13	19.1		11	15.7				
Controlled Drinking	28	14.3		10	17.2		9	13.2		9	12.9				
Assessment at baseline															
TLFB	241	99.6		79	100		83	98.8		79	100				1.00
Questionnaire	241	99.6		79	100		83	98.8		79	100				1.00
Assessment at 3-month FU															
TLFB	173	71.5		57	72.2		58	69.0		58	73.4				.816
HDL-Questionnaire	142	58.7		45	57		49	57.1		49	62.0				.762
Short Interview	169	69.8		55	69.6		57	67.9		57	72.2				.836

Note. Motivation and Confidence as reported in this table were assessed with single questions (31, 32) to which patients answered on a 11-point Likert Scale. Therapy goals were assessed as multiple-choice questions. The rows captioned Assessment at baseline and Assessment at 3-month-FU indicate the amount of data available at the respective time point.

Abbreviations: ASPD: antisocial personality disorder ; ASRS: Adult ADHD Self-Report Scale ; BAI: Beck Anxiety Inventory ; BDI: Beck's Depression Inventory ; BISBAS: Behavioral Inhibition System/Behavioral Approach System Scale ; HDL: Health and daily-living form; FU: 3-Month Follow-up; I8: Scale for Impulsive Behavior ; NISS: Need Inventory of Sensation Seeking ; SCI: Stress and Coping Inventory

2.2 Additional results regarding training effects on drinking outcomes: linear regression models

2.2.1 Sensitivity analysis: Regression analysis of the training effects on PDA using alternative assumptions regarding missing data.

As described in section 1.7, we used multiple imputations to deal with missing outcomes in the main analyses of PDA and PHDD. For comparison, we also estimated PDA Models under the two alternative assumptions MNAR (last-observation-carried-forward) and MCAR (complete-case). Interpretation of these analyses is limited by two facts: I) In our data set, MAR, and not MCAR or MNAR, was the assumption with the best support. II) In a comparison of methods to deal with missing data in RCTs on AUD, it was shown that complete-case analyses and last-observation-carried-forward analysis were shown to provide more biased results than analyses using multiple imputations (37).
In eTable 2 the estimates of these are presented. Note that the effects of standard and improved Alc-IT point in the same direction as in the main analysis. The analyses under MAR using multiple imputations were generally more efficient as can be seen from the shorter confidence intervals and lower p-Values. The reason for this may be in both cases a reduced statistical power: a) Last-observation-carried-forward leads to higher intercorrelation between the longitudinal measurements. Thus, inflation of the observed/calculated correlation reduces the information in the data and thereby results in diminished statistical power; b) The reduced sample size of the complete-case subset leads to reduced statistical efficiency of effect estimates while increasing the potential for bias.

eTable 2: Analyses of PDA at 3-month follow-up under a MCAR and MNAR assumption

			MNAR				**MCAR**	
Effect	Est	*SE*	95% CI	*p*	Est	*SE*	95% CI	*p*
Intercept	57.50	4.59	48.46-66.54	< .001	81.68	3.91	73.96-89.41	< .001
PDA Baseline	0.28	0.08	0.12-0.44	< .001	0.05	0.07	-0.09-0.19	.458
Standard Alc-IT	6.75	5.88	-4.82-18.33	.0252	3.04	4.92	-6.67-12.76	.537
Improved Alc-IT	11.69	5.88	0.11-23.27	.048	9.87	4.90	0.20-19.53	.045
N	242 Patient				173 Patient			

Abbreviations: Alc-IT: alcohol-specific inhibition training; CI: confidence interval; Est: Estimate; MNAR: Regression model based on last-observation-carried-forward data (under the assumption of missing not at random); MCAR: Regression model based on complete-case data (under the assumption of missing completely at random); PDA: Percentage of days abstinent; SE: standard error

2.2.2 Directly testing the effects of standard and improved Alc-IT on PDA against each other

eTable 3 Regression models directly comparing two training interventions against each other

	Control vs standard Alc-IT			**Control vs improved Alc-IT**			**Standard vs improved Alc-IT**		
Predictors	*Est*	*CI*	*p*	*Est*	*CI*	*p*	*Est*	*CI*	*p*
Intercept	73.21	65.48 – 80.95	**<0.001**	78.27	71.88 – 84.66	**<0.001**	68.43	61.42 – 75.44	**<0.001**
PDA Baseline	0.07	-0.09 – 0.23	0.370	-0.13	-0.27 – 0.00	0.055	0.13	-0.02 – 0.28	0.083
Training Intervention	-3.34	-12.67 – 6.00	0.481	11.42	3.73 – 19.11	**0.004**	14.84	6.24 – 23.44	**0.001**
Observations	162			158			162		
R^2 / adj. R^2	.008 / -.004			.074 / .062			.084 / .073		

Abbreviations: adj: adjusted; Alc-IT: alcohol-specific inhibition training:; CI: 95% confidence interval; Est: Estimate; MNAR: Regression model based on last-observation-carried-forward data (under the assumption of missing not at random); MCAR: Regression model based on complete-case data (under the assumption of missing completely at random); PDA: Percentage of abstinent days abstinent; SE: standard error

2.3 Additional Results, complementary analysis: Hierarchical linear models testing training effects on PDA and PHDD

To provide a complementary analysis, which takes the repeated measurement and alternative estimates (indicators) of the outcomes PDA and PHDD into account, training effects were additionally assessed in hierarchical linear models. A series of hierarchical linear models nesting timepoints and different PDA/PHDD indicators within patients were developed. Two indicators of percentage of days abstinent at baseline (derived from the TLFB and the questionnaire) and three indicators of percentage of days abstinent at 3-month follow-up (derived from TLFB, HDL questionnaire, and short interview) were integrated into the analyses. While interpreting the HLM-results, one needs to recognise that the integration of assessments from 3 different types of measurement entails the problem that those measures might plausibly yield diverging results. On the other hand, one might argue that a model integrating all available measures adds valuable information in terms of the robustness of the analysed effects. In case of the data at hand, drinking outcome assessed with the HDL and TLFB were found to be strongly positively correlated, (PDA: $r_{(140)} = .815$, $p < .001$; PHDD: $r_{(140)} = .913$, $p < .001$), the same was true for measure obtained with the short interview and the TLFB (PDA: $r_{(167)} = .895$, $p < .001$; PHDD: $r_{(167)} = .835$, $p < .001$). Further, the values obtained with the short interview and the HDL-questionnaire were also associated (PDA: $r_{(136)} = .837$, $p < .001$; PHDD: $r_{(136)} = .893$, $p < .001$).

Of the total sample of 242, 241 (99.5%) at baseline and 173 (71.5%) at 3-month follow-up provided data indicating the number of days spent in a protected environment as well as TLFB data. Of these 173 provided additional indicators for drinking outcome data, 169 (HDL questionnaire), respectively 142 (short interview). For the conservative intention-to-treat analyses, all 242 subjects were retained in the hierarchical linear model (HLM) analyses.

2.3.1 Hierarchical Linear Models Structure

With fully unconditional two-level models, the necessity of nesting the data within patients was tested.

Next, fully unconditional three-level models were computed in which the dependent variables were nested within patients nested within the three study sites. Comparing the respective two and three-level models indicated that including the study site did not add sufficient information to justify the more complex three-level models (i.e., resulting in a singular fit).

Next, *time* was included as a predictor to the two-level models. The resulting models showed a significant increase in explained variance compared to the fully unconditional model.

Next, *intervention* was included as a predictor of the intercept to test for unintended group differences at the baseline. However, no significant effect of intervention on the intercept was observed, and the model fit was not significantly improved by including intervention as a predictor of the intercept. Thus, there was no indication of unintended baseline differences between groups. This finding, as well as the randomization procedure, indicates that intervention as a predictor of the intercept can be omitted from the models. However, to take a conservative approach to analysis, we examined both models (i.e., models incorporating an effect of the intervention on the intercept and models not incorporating an effect of intervention on the intercept).

Next, an interaction term between time *and* intervention was included in the models. This term reflects the effect of the intervention on the change from baseline to follow-up measurement.

Finally, we examined covariates in the models that could potentially act as confounders that could bias the estimates. Taking into account possible multicollinearity we considered: *age, gender, daytime of the*

training; days in residential treatment; and *pharmacotherapy.* The models incorporating *days in residential treatment, and pharmacotherapy* showed an effect significantly different from zero on the change in explained variance. These two variables were thus included in the models.

2.3.2 HLM estimates of the primary outcome PDA

To estimate the effect of Alc-IT on the change of PDA over time, the following equations were implemented as HLMs:

Level 1

$$PDA_{ij} = \beta_{0j} + \beta_{1j} * \textit{Time}_{ij} + r_{ij} \quad (1.1)$$

Level 2

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} * \textit{Alc-IT}\ standard + \gamma_{02} * \textit{Alc-IT}\ improved + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} * \textit{Alc-IT}\ standard + \gamma_{12} * \textit{Alc-IT}\ improved \end{aligned} \quad (1.2)$$

which leads to the following combined HLM:

$$\begin{aligned} PDA_{ij} = \gamma_{00} &+ \gamma_{01} * \textit{Alc-IT}\ standard + \gamma_{02} * \textit{Alc-IT}\ improved + \gamma_{10}(\textit{Time}_i) \\ &+ \gamma_{11} * \textit{Alc-IT}\ standard + \gamma_{12} * \textit{Alc-IT}\ improved + u_{0j} + r_{ij} \end{aligned} \quad (2)$$

Equation 2 shows that the PDA of a specific $patient\ (j)$ at a certain point in *time* (i) is a function of the intercept of *PDA* (γ_{00}), if *Time* $= Baseline$ and $intervention =$ *control training*), the effect of Alc-IT on the intercept $(\gamma_{01} + \gamma_{02})$, the effect of *time* (i.e., the change from baseline to 3-month follow-up, γ_{10}), the interaction effect of *time* and Alc-IT $(\gamma_{11} + \gamma_{12})$, the patient's *PDA* deviation from the sample's mean intercept (u_{0j}), and the patient's (j) residual error at *time* $= i$ (r_{ij}). (Note that the effect of the covariates is not included in the equation.)

eTable 4a shows the estimated model parameters for the models estimating the effects of Alc-IT without an effect of Alc-IT on the intercept (M.PDA) and with an effect of Alc-IT on the intercept (M.PDA.intercept). Both include the covariates *days in residential treatment* and *Pharmacotherapy*.

eTable 4: Estimates of M.PDA, M.PDA.intercept, M.PHDD, and M.PHDD.intercept

A

	M.PDA			M.PDA.intercept		
Effect	**Estimate**	**CI**	***p***	**Estimate**	**CI**	***p***
Intercept	23.23†	11.58 – 34.88	< .001	23.23†	11.58 – 34.88	< .001
Time	35.44†	24.01 – 46.86	< .001	35.44†	24.01 – 46.86	< .001
Alc-IT std.				2.3	-5.22 – 9.81	.549
Alc-IT imp.				2.13	-5.41 – 9.68	.579
Pharmacotherapy	34.44	-17.62 – 86.50	.194	34.44	-17.62 – 86.50	.194
Days in residential treatment	-0.03	-0.15 – 0.10	.691	-0.03	-0.15 – 0.10	.691
Time*Alc-IT.std	-2.44†	-10.12 – 5.24	.534	-4.74†	-11.92 – 2.45	.196
Time*Alc-IT.imp	10.46†	2.72 – 18.21	.008	8.33†	1.11 – 15.55	.024
Time*Pharmacotherapy	45.7†	-9.39 – 100.79	.104	45.7†	-9.39 – 100.79	.104
Time*Days in residential treatment	0.32†	0.19 – 0.45	< .001	0.32†	0.19 – 0.45	< .001
Random Effect						
σ^2	538.69			490.51		
τ_{00}	310.32 $_{Patient}$			310.32 $_{Patient}$		
ICC	0.4			0.32		
N	242 $_{Patient}$			242 $_{Patient}$		
Observations	966			966		
Marginal R^2 / Conditional R^2	0.58 / 0.75			0.58 / 0.75		

B

	M.PHDD			M.PHDD.intercept		
Effect	**Estimate**	**CI**	***p***	**Estimate**	**CI**	***p***
Intercept	73.07†	61.25 – 84.89	< .001	73.07†	61.25 – 84.89	< .001
Time	-42.15†	-53.77 – -30.53	< .001	-42.15	-53.77 – -30.53	< .001
Alc-IT std.				-3.67	-11.29 – 3.95	.345
Alc-IT imp.				-2.46	-10.12 – 5.20	.529
Pharmacotherapy	-24.97	-77.73 – 27.80	.353	-24.97	-77.73 – 27.80	.353
Days in residential treatment	-0.01	-0.14 – 0.12	.908	-0.01	-0.14 – 0.12	.908
Time*Alc-IT.std	0.59†	-7.20 – 8.38	.882	4.26†	-2.99 – 11.51	.249
Time*Alc-IT.imp	-8.2 †	-16.06 – -0.34	.041	-5.74 †	-13.03 – 1.55	.123
Time*Pharmacotherapy	-12.66†	-68.33 – 43.01	.656	12.66†	-68.33 – 43.01	.656
Time*Days in residential treatment	-0.24†	-0.37 – -0.10	<.001	-0.24†	-0.37 – -0.10	<.001
Random Effect						
σ^2	477.23			477.23		
τ_{00}	321.10 $_{Patient}$			321.10 $_{Patient}$		
ICC	0.4			0.4		
N	242 $_{Patient}$			242 $_{Patient}$		
Observations	964			964		
Marginal R^2 / Conditional R^2	0.56 / 0.73			0.56 / 0.73		

Note. The upper left part of the table (i.e. the part reporting on the model M.PDA) was already shown in table 3 of the main paper and is repeated here only to allow for direct comparison of the model estimates with and without the inclusion of the intercept. Abbreviations: Alc-IT.std, standard alcohol-specific inhibition training with an equiprobable ratio of Go and NoGo trials; Alc-IT.imp, improved alcohol-specific inhibition training with a Go/NoGo ratio of 75/25; CI, 95% confidence interval; p, probability of the estimated parameter value (or of more extreme parameter values), given that the true coefficient is zero; Estimates, estimated regression coefficients; σ^2, random effect variance; τ_{00} , random-intercept-variance;; ICC, Intraclass-correlation coefficient; N, sample size; Marginal R^2, explained variance by fixed effects; Conditional R^2, explained variance by the entire model, that is, both fixed effects and random effects. †: Significant improvement in fit over previous model at $p < .05$

Directly testing the effects of standard and improved Alc-IT on PDA against each other: To estimate the difference between the effects of standard and improved Alc-IT, we directly compared the standard Alc-IT with the improved Alc-IT in an additional model. This model yielded a significant effect indicating that, compared to participants in standard Alc-IT, participants in improved Alc-IT reported a stronger increase in PDA from baseline to 3-month follow-up (γ_{11} = 11.46, SE = 3.61, 95% CI [4.35, 18.55], $t_{(df=545)}$ = 3.17, p = .002, f^2 = 0.12).

Does endogenous cortisol moderate the effect of Alc-IT on PDA? The present study was also designed in a way that allowed the investigation of the moderating effects of endogenous cortisol, which is known to enhance learning, on potential training effects (see, (6)). Therefore, we computed additional models incorporating such moderating effects. However, when daytime of training was included as a predictor in M.PDA, the model yielded no effect of daytime of training on changes in PDA significantly different from zero (γ = 0.98, *SE* = 7.37, 95% CI [-13.48, 15.44], $t_{(df=537)}$ = 0.13, p = .894).

2.3.3 HLM estimates of the outcome PHDD

To estimate the effect of standard Alc-IT and improved Alc-IT on the change in PHDD, Equation 2 was adjusted accordingly. eTable 4B shows the estimated model parameters.

2.4 Training intervention effects on time to first drink

Fitted Kaplan Meyer curves showed that the group receiving improved Alc-IT showed the highest probability of abstinence over the entire 90-day period, although no significant differences were observed between the three training intervention groups (χ2(2) = 2.47, p = .300). Comparing the control training with standard Alc-IT, the contrast was not significant, χ2(1) = 0.09, p = .769, as was the contrast between the control training and improved Alc-IT, χ2(1) = 2.34, p = .126. Likewise, no significant contrast was observed comparing the standard Alc-IT with improved Alc-IT, χ2(1) = 1.43, p = .233.

2.5 Additional results concerning the Go-NoGo-task

Errors of commission from the Go-NoGo-task were analyzed non-parametrically using Anova-type non-parametric statistics (ATS) in a three-factorial design including the within-factors time (pre, post) and picture type (alcohol, neutral) as well as the between-factor training group (standard Alc-IT, improved Alc-IT, control). This overall analysis yielded a significant time by training group by picture type interaction ($ATS_{(df=2)}$ = 11.07, p = .004), as well as significant main effects of time ($ATS_{(df=1)}$ = 10.91, p = .001) and picture type ($ATS_{(df=1)}$ = 120.8, p < .001) and a significant interaction between training group and picture type ($ATS_{(df=2)}$ = 4.61, p = .010). All other main effects and interactions were not significant (all ATS < 1.35, all p > .2). To better understand the three-way interaction between time, training group, and picture type, the follow-up analyses investigated the effects of the factors picture type and time separately in each training group.

For improved Alc-IT, this yielded a significant time by picture type interaction ($ATS_{(df=1)}$ = 9.9, p = .002). The main effect of picture type was also significant ($ATS_{(df=1)}$ = 79.05, p < .001), but no main effect of time was observed ($ATS_{(df=1)}$ = 1.36, p = .244). In order to understand the significant tow way interaction between time and picture type in this group, a second follow-up analysis investigated the effect of time separately for alcohol-related and neutral errors of commission and indicated that alcohol-related errors of commission decreased stronger from pre to post-training ($ATS_{(df=1)}$ = 5.00, p = .025) than neutral errors of commission ($ATS_{(df=1)}$ = 0.01, p = .913, eFigure 4 (middle panel)).

For standard Alc-IT and the control group, no such interaction was observed (both $ATS_{(df=1)}$ < 1.95, both p > .160). For standard Alc-IT, analyses yielded a main effect of picture type ($ATS_{(df=1)}$ = 26.47, p < .001)

and a main effect of time ($ATS_{(df=1)}$ = 8.30, p = .004), for the control group a main effect of picture type ($ATS_{(df=1)}$ = 25.48, p < .001) and only a trend towards a main effect of time ($ATS_{(df=1)}$ = 3.15, p = .076) was observed. Descriptive values for errors of commission, errors of omission and reaction times are displayed in eTable 5.

2.5.1 Errors of commission as a mediator of the effect of Alc-IT on the change in PDA

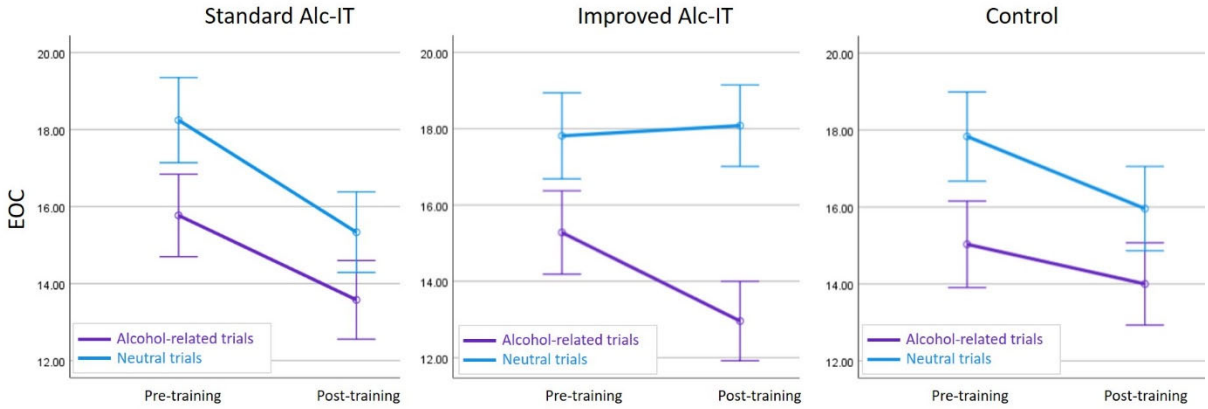
In a mediation analysis, we tested the hypothesis that the effect of Alc-IT on the change in PDA is mediated by an improvement in the errors of commission (Go-NoGo-task). There was no indication found that the indirect effect of Alc-IT on the change in PDA mediated by is different from zero (b = -0.08, *SE* = 0.326; Z = -0.250, *p* = .803). In this model the effect of Alc-IT on the errors of commission (b = 0.422, *SE* = 1.63, Z = -0.26, *p* = .795) and the effect of the errors of commission on PDA (b = 0.073, *SE* = 0.169, Z = -0.433, *p* = .665) were both not significantly different from zero.

eTable 5: Errors of commission and reaction times during the Go-NoGo task

	Total Sample			Standard Alc-IT			Improved Alc-IT			Control		
	M	Med	SD	M	Med	SD	M	Med	SD	M	Med	SD
EOCs												
Pre-training												
EOC (Alc)	15.44	14	9.46	15.93	14	10.25	15.25	14	9.17	15.13	14	8.97
EOC (Neu)	17.87	17	9.67	18.34	18	9.68	17.74	17	9.47	17.49	15	9.95
Post-training												
EOC (Alc)	13.50	12	8.99	13.58	11	8.95	12.96	10	8.53	14.00	12	9.58
EOC (Neu)	16.45	15	9.28	15.33	13	9.40	18.08	17	9.50	15.96	14	8.78
RTs												
Pre-training												
RT (Go Alc)	424.53	424.45	66.59	427.21	428.61	63.41	421.50	412.69	72.28	424.71	425.82	64.73
RT (Go Neu)	421.04	420.82	65.33	423.35	427.28	59.49	416.42	407.89	71.25	423.15	422.30	65.69
Post-training												
RT (Go Alc)	391.96	382.78	63.80	398.56	390.87	58.09	384.83	379.38	66.10	392.24	381.11	67.33
RT (Go Neu)	388.46	382.43	62.16	397.82	387.46	56.35	376.59	371.94	63.27	390.72	387.72	65.82
EOOs												
Pre-training												
EOO (Alc)	10.34	4	21.32	10.06	5	17.55	12.25	5	28.71	8.77	4	15.95
EOO (Neu)	9.36	3	20.30	8.10	3	14.22	11.32	4	27.84	8.76	4	16.83
Post-training												
EOO (Alc)	6.95	3	17.87	9.86	3	26.98	5.43	3	8.14	5.37	2	11.48
EOO (Neu)	5.92	2	17.34	8.73	2	26.24	3.95	2	7.62	4.92	1	11.16

Note. Abbreviations: EOC, errors of commission (in NoGo trials); EOO, errors of omission (in Go trials); RT, Reaction times; Alc, alcohol-related trials; Neu, neutral trials; M, mean; Med, median; SD, standard deviation; Standard Alc-IT, standard alcohol-specific inhibition training with an equiprobable ratio of Go and NoGo cues; improved Alc-IT, improved alcohol-specific inhibition training with a Go/NoGo ratio of 75/25; Control, control training. Pre-training, assessment conducted 1-4 days prior to the first training session. Post-training, assessment conducted 1-4 days after the last training session

eFigure 4: Errors of commission during the Go-NoGo-task



Note. Errors of commission are shown at pre- and post-training assessments for each picture type and training group. Error bars represent 1 standard error. Abbreviations: EOC, errors of commission; Standard Alc-IT, standard alcohol-specific inhibition training with an equiprobable ratio of Go and NoGo cues; improved Alc-IT, improved alcohol-specific inhibition training with a Go/NoGo ratio of 75/25; Control, control training. Pre-training, assessment conducted one to four days prior to the first training session. Post-training, assessment conducted one to four days after the last training session.

2.6 Additional results concerning a change in craving induced by the training intervention

A change in subjective craving from baseline to 3-months follow-up was listed as a secondary outcome in the trial registration and was assessed with the German version of the obsessive compulsive drinking scale (OCDS, 4). A repeated-measures ANOVA with the factors time (baseline, 3-month follow-up) and training group (standard Alc-IT, improved Alc-IT, and control) and the overall score of the OCDS was computed. This analysis yielded a significant effect of time ($F_{df=1}$ = 458.13, p < 0.001, η^2 =.78) indicating that all three groups reported decreased training at follow-up measurement, but no time × training group interaction ($F_{df=2}$ = 2.74, p = 0.07, η^2 = .04). Thus, the analysis does not support the assumption that the training influenced the subjectively experienced craving.

3 References

1. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.
2. Batschelet HM, Tschuemperlin RM, Moggi F, Soravia LM, Koenig T, Pfeifer P, et al. Neurophysiological correlates of alcohol-specific inhibition in alcohol use disorder and its association with craving and relapse. Clin Neurophysiol. 2021;132(6):1290-301.
3. Bott KE, Rumpf H-J, Bischof G, Meyer C, Hapke U, John U. Alcoholabstinenz-Selbstwirksamkeitsfragebogen; Deutsche Version (AASE-G); deutsche Version der Alcohol abstinence Self-efficacy-Scale, German Version. In: Glöckner-Rist A, Rist F, Küfner H, editors. Elektronisches Handbuch zu Erhebungsinstrumenten im Suchtbereich (EHES). 3 ed. Mannheim: Zentrum für Umfragen, Methoden und Analysen; 2003.
4. Mann K, Ackermann K. Die OCDS-G: Psychometrische Kennwerte der deutschen Version der OBsessive and Complusive Drinking Scale. Sucht. 2000;46:90-100.
5. Houben K, Havermans RC, Nederkoorn C, Jansen A. Beer a no-go: learning to stop responding to alcohol cues reduces alcohol intake via reduced affective associations rather than increased response inhibition. Addiction. 2012;107(7):1280-7.
6. Tschuemperlin RM, Stein M, Batschelet HM, Moggi F, Soravia LM. Learning to resist the urge: a double-blind, randomized controlled trial investigating alcohol-specific inhibition training in abstinent patients with alcohol use disorder. Trials. 2019;20(1):402.
7. Houben K, Nederkoorn C, Wiers RW, Jansen A. Resisting temptation: decreasing alcohol-related affect and drinking behavior by training response inhibition. Drug Alcohol Depend. 2011;116(1-3):132-6.
8. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one decade on. Trends Cogn Sci. 2014;18(4):177-85.
9. Smith JL, Mattick RP, Jamadar SD, Iredale JM. Deficits in behavioural inhibition in substance abuse and addiction: a meta-analysis. Drug Alcohol Depend. 2014;145:1-33.
10. Noguchi K, Gel YR, Brunner E, Konietschke F. nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. J Stat Softw. 2012;50(12):1-23.
11. Shelton J, Kumar GP. Comparison between Auditory and Visual Simple Reaction Times Neuroscience and Medicine. 2010;1(1):30-2.
12. Egenolf Y, Stein M, Koenig T, Grosse Holtforth M, Dierks T, Caspar F. Tracking the implicit self using event-related potentials. Cognitive, affective & behavioral neuroscience. 2013;13(4):885-99.
13. Tschuemperlin RM, Batschelet HM, Moggi F, Koenig T, Roesner S, Keller A, et al. The Neurophysiology of Implicit Alcohol Associations in Recently Abstinent Patients With Alcohol Use Disorder: An Event-Related Potential Study Considering Gender Effects. Alcohol Clin Exp Res. 2020.
14. Greenwald AG, Nosek BA, Banaji MR. Understanding and using the implicit association test: I. An improved scoring algorithm. Journal of personality and social psychology. 2003;85(2):197-216.
15. Fey W, Moggi F, Rohde KB, Michel C, Seitz A, Stein M. Development of stimulus material for research in alcohol use disorders. Int J Methods Psychiatr Res. 2017;26(1):e1527.
16. Verbruggen F, Aron AR, Band GPH, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. Elife. 2019;8.
17. Wittchen HU, Pfister H. Instruktionsmanual zur Durchführung von DIA-X-Interviews. Frankfurt: Swets Test Services; 1997.
18. Wechsler H, Davenport A, Dowdall G, Moeykens B, Castillo S. Health and behavioral consequences of binge drinking in college. A national survey of students at 140 campuses. JAMA. 1994;272(21):1672-7.
19. Barbor TF, De la Fuente JR, Saunders J, grant M. Audit. The alcohol Use Disorders Identification Test. Guidelines for Use in Primary Health Care. Geneva: World Health Organization; 1992.

20. Berman AH, Bergman H, Palmstierna T, Schlyter F. Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a swedish population sample. Eur Addict Res. 2005;11(1):22–31.
21. Miller WR, Tonnigan JS. Assessing Drinkers motivation for change: The stages of change readiness and treatment eagerness scale (SOCRATES). Psychology of Addictive Behaviors. 1996;10:81-9.
22. Franke GH. Brief Symptom Inventory von L.R. Derogatis - Deutsches Manual. [Brief symptom INventory of L.R. Derogatis - German Manual]. Göttingen: Beltz; 2000.
23. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59(20):22-33.
24. Lauth GW, Minsel W-R. KATE. Kölner ADHS-Test für Erwachsene. Göttingen: Hoegrefe; 2014.
25. Beck AT, Steer RA. Beck Anxiety Inventory Manual San Antonio The Psychological Corporation; 1993.
26. Beck AT, A. SR, Brown GK. BDI-II Beck Depressions-Inventar 2. Auflage. Frankfurt/Main: Harcourt; 2006.
27. Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. JPSP. 1994;67(2):319-33.
28. Kovaleva A, Beierlein C, Kemper CJ, Rammstedt B. Eine Kurzskala zur Messung von Impulsivität nach dem UPPS-Ansatz: Die Skala Impulsives-Verhalten-8 (I-8). Working Papers. 2012;20:5-31.
29. Roth M, Hammelstein P. The Need Inventory of Sensation Seeking (NISS). EJPA. 2012;28(1):11-8.
30. Satow L. Stress- und Coping-Inventar (SCI): Test- und Skalendokumentation 2012 [updated 03.07.2018. Available from: http://www.drsatow.de].
31. Ludwig F, Tadayon-Manssuri E, Strik W, Moggi F. Self-efficacy as a predictor of outcome after residential treatment programs for alcohol dependence: simply ask the patient one question! Alcohol Clin Exp Res. 2013;37(4):663-7.
32. Foster S, Held L, Estevez N, Gmel G, Mohler-Kuo M. Liberal alcohol legislation: does it amplify the effects among Swiss men of person-related risk factors on heavy alcohol use? Addiction. 2015;110(11):1746-56.
33. Sobell LC, Sobell MB. Alcohol Timeline Followback users' manual. Toronto, CA: Addiction Research Foundation; 1995.
34. Moos RH, Cronkite RH, Finney JW. Health and Daily Living Form Manual. Palo Alto, CA: Mind Garden; 1990.
35. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017;17(1):162.
36. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.
37. Hallgren KA, Witkiewitz K. Missing data in alcohol clinical trials: a comparison of methods. Alcohol Clin Exp Res. 2013;37(12):2152-60.
38. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mxed-effects models using lme4. arXiv preprint. 2014;arXiv:1406.5823.
39. Wickham H. gplot2: elegant graphics for data analysis: springer; 2016.
40. Kumar D, Klefsjö B. Proportional Hazards Model: A Review. Reliability Engineering & System Safety. 1994;44(2):177-88.