Data supplement for Arad et al., Attention Bias Modification Treatment Versus a Selective Serotonin Reuptake Inhibitor Or Waiting List Control for Social Anxiety Disorder: A Randomized Clinical Trial. Am J Psychiatry (doi: 10.1176/appi.ajp.20220533)

Statistical Analysis Plan

Data preparation

All data was entered into SPSS, to create one database containing all relevant variables and time-points for each participant (available in the OSF repository, Identifier: DOI 10.17605/OSF.IO/MYA75). Participants were represented by the "ID" variable. Group allocation was represented by the "group" variable (0=Patients who were assessed and excluded or not randomized, 1=randomized to GC-MRT, 2=randomized to SSRI, 3=randomized to waitlist). Completion of the study was represented by the "completion" variable (0= not randomized, 1=completed all stages, 2=dropped out). "Adherence" represented the number of sessions attended by the patient (for GC-MRT the full course included 10 sessions, for SSRI 4 sessions).

Where relevant, questionnaires were entered into separate SPSS files with questions as variables, to compute Cronbach's alpha scores. Then, total scores were summed up and entered into the final database for statistical analysis. CEQ scores were analyzed as recommended by Devilly et al (2000)[1], using z-scores as the analyzed variable.

Clinically Significant Change (CSC) and Reliable Change (RC)[2] cutoffs were determined based on the test-retest reliability data from Baker et al[3] and pre-treatment LSAS scores from the authors' data from previous trials (N=169). The LSAS cutoff was set at 46.29 for CSC and the RC cutoff was set at 1.96.

Percent dwell time on threat was calculated using Excel. First, a fixation report was extracted from the EyeLink Data Viewer software (version 2.5.0.14; SR Research Ltd., Mississauga, Ontario, Canada) to Excel. Ares of interest (AOI) type (i.e threat/neutral) was determined according to the current fixation interest area ID and the pre-prepared maps of the presented face matrices detailing stimuli type and location. First fixations on the fixation cross (current fixation index=1) and fixations shorter than 100ms (current fixation duration<100) were flagged and left out of the analysis. Then, using a PIVOT table, dwell times (current fixation duration) were averaged separately for the threat and neutral AOIs, and separately for each time point, participant and matrix. Percent dwell time on threat was calculated as the proportion of time fixating on the threat AOI relative to the total time fixating on faces (threat+neutral AOIs). Cronbach's alpha was calculated between the 30 matrices, separately for each time point, treating each dwell time % on threat in each matrix presentation as an item. For statistical analysis, percent dwell time on threat scores were averaged per participant per time-point and entered in the final database.

Data Analysis

All analyses were conducted using IBM SPSS, version 28.0.1.0.

Continuous outcome measures with more than one time-point were LSAS (pre-, mid-, post-treatment), SPIN (pre-treatment, weeks 2, 4, 6, 8, 10, post-treatment), PHQ-9 (pre-, post-treatment) and percent dwell time on threat (pre-, mid, post-treatment). Outcomes for these variables were analyzed using generalized estimating equations (GEE) as an intention-to-treat approach accommodating for missing data and considering correlations between repeated-measures via estimated marginal means and based on data from all randomized participants. To represent within-subject dependencies, an unstructured correlation matrix was chosen. For each outcome (LSAS, SPIN, PHQ, dwell-time) we first applied a full factorial model containing the effects of time, group, and their interaction as predictors. Pairwise comparisons were chosen for follow-up analysis.

For the initial analysis, the time-by-group interaction term represented differential clinical or attentional effects between the three treatment conditions. Significant interaction effects were followed-up with the pairwise comparisons of each of the active treatment groups with the control group and with each other, separately for each time point (e.g, GC-MRT-SSRI, SSRI-Control, GC-MRT-Control, at pre-, mid- and post-treatment, separately). For the primary outcome (LSAS) and for attentional dwell time on threat, 3 time points were considered (pre-, mid-, and post-treatment); 7 time points (pre-treatment, weeks 2, 4, 6, 8, 10, and post-treatment were considered for the secondary outcome (SPIN); 2 time points (pre- and post-treatment) were considered for depression (PHQ-9). For each pairwise comparison, and based on the estimated means and SDs, Cohen's d was also calculated along with the 95% confidence interval for this effect. For dwell-time on threat, an additional GEE analysis with Time(3) as the only predictor was conducted separately for each group to assess a unique mechanism of attentional training.

Chi-square tests were used to compare the number of patients displaying CSC and RC at mid- and post-treatment in the different groups. To tackle the issue of missing data, patients who dropped-out were considered as not displaying RC or CSC for this analysis.

CGI-I scores, rated by clinicians, represented clinical improvement. Due to small variance in this scale, scores were converted to binary scores, with the cutoff being CGI-I=2 (much improved). Then, the number of patients rated 'much' (CGI=2) or 'very much' (CGI=1) improved at post-treatment was compared to the number of patients rated as 'minimally' (CGI=3) or 'not' (CGI>4) improved across groups. Patients who dropped-out were rated as not improved.

To examine differences in expectancy and credibility (CEQ) at pre-treatment, and satisfaction with treatment (CSQ) at post-treatment, GEE analyses were conducted contrasting the two active treatment groups, with group (2) as the sole predictor.

Within the GC-MRT group, session-to-session changes in dwell time on threat were estimated using GEE, with time (10) as the sole predictor. To examine generalization of training to a new set of faces and in the absence of music reinforcement, a Pearson correlation between reduction in dwell time on threat from training sessions 1 to 10 and reduction in dwell time on threat from baseline to post-treatment was used.

Missing data

All missing data points are detailed below. Missing data not resulting from drop-out is explained in footnotes.

	GC-MRT		SSRI		Control	
	Valid	Missing	Valid	Missing	Valid	Missing
LSAS						
Baseline	35	0	35	0	35	0
Mid-treatment	34	1	26	9	35	0
Post-treatment	33	2	25	10	32	3
SPIN						
Baseline	35	0	35	0	35	0
Week 2	34	1	26	9	35	0
Week 4	34	1	25	10[1]	33	2[2]
Mid-treatment	34	1	25	10[3]	34	1[4]
Week 8	34	1	26	9	33	2[5]
Week 10	32	3[6]	24	11[7]	34	1
Post-treatment	33	2	25	10	31	4[8]
PHQ						
Baseline	35	0	35	0	35	0
Post-treatment	33	2	25	10	30	5[9]
CGI						
Mid-treatment	33	2[10]	26	9	34	1[11]
Post-treatment	33	2	25	10	32	3
CEQ	34	1[12]	27	8[13]	--	--
CSQ	33	2	25	10	--	--
Dwell-time- measurement						
Baseline	35	0	35	0	35	0
Mid-treatment	33	2[14]	25	10[15]	34	1[16]
Post-treatment	33	2	25	10	32	3
Dwell-time- GC-MRT training						
session1	35	0	--	--	--	--
session2	35	0	--	--	--	--

[1] 8153- did not fill-out questionnaire

[2] 8068, 8117- did not fill-out questionnaire

[3] 8157- did not fill-out questionnaire

[4] 8117- did not fill-out questionnaire

[5] 8082- did not fill-out questionnaire

[6] 8074- did not fill-out questionnaire

[7] 8143- did not fill-out questionnaire

[8] 8221- did not fill-out questionnaire

[9] 8212, 8221- did not fill-out questionnaire

[10] 8262- did not fill-out questionnaire

[11] 8117- did not fill-out questionnaire

[12] 8038- did not fill-out questionnaire

[13] 8014, 8064, 8125, 8141, 8164, 8166, 8195, 8214- did not fill-out questionnaire

[14] 8028- damaged file

[15] 8153- performed clinical assessment without the cognitive task

[16] 8117- performed clinical assessment without the cognitive task

session3	34	1[17]	--	--	--	--
session4	34	1	--	--	--	--
session5	34	1	--	--	--	--
session6	34	1	--	--	--	--
session7	34	1	--	--	--	--
session8	33	2[18]	--	--	--	--
session9	33	2[19]	--	--	--	--
session10	30	5[20]	--	--	--	--

Completers Analysis for Categorical Variables

Clinically Significant Change and Reliable Change

Of the patients in the GC-MRT, SSRI, and the control groups, 39.4%, 40%, and 6.3% showed clinically significant change following treatment, respectively ($\chi^2(2)=11.47$, p=0.003). Follow-up analyses indicated higher frequencies of patients with clinically significant change in the GC-MRT and SSRI groups relative to the control group (χ^2s(1)=10.05 and 9.61, ps=0.002, respectively), with no difference between the two active groups ($\chi^2(1)=0.002$, p=0.96). Reliable change was noted in 51.5% of the patients in the GC-MRT group, in 68% the SSRI groups, and in 15.6% in the control group ($\chi^2(2)=17.10$, p<0.001). Higher frequencies of reliable change were noted in the GC-MRT and SSRI relative to the control group separately (χ^2s(1)=9.35 and 16.25 p=0.002 and p<0.001, respectively), with no significant difference between the two active groups ($\chi^2(1)=1.59$, p=0.21).

CGI

Improvement was rated by clinicians as 'much' or 'very much' improved at post-treatment in 42.4% of patients in the GC-MRT group, 56% in the SSRI group, and 9.4% in the control group (CGI, $\chi^2(2)=14.98$, p<0.001), with higher percentages of improved patients in each of the GC-MRT and SSRI groups relative to the control group (χ^2s(1)=9.20 and 14.58, ps=0.002 and p<0.001, respectively), and no significant differences between the active treatments (χ^2 (1)=1.05, p=0.306).

---

[17] 8198- damaged file
[18] 8001- damaged file
[19] 8152- did not perform training session
[20] 8074, 8149, 8165- did not perform training session

Missingness at Random Analysis

Due to the differential drop-out between groups, baseline measures of completers and drop-out groups were compared using independent samples Welch's t-tests and Chi-squared tests, with two-sided alpha ≤0.05. The variables tested were age, baseline social anxiety and depression levels. No differences were noted between treatment completers and non-completers (see Table 1S below). In addition, Little's test of missing completely at random was conducted using the same baseline demographic and clinical measures to examine the pattern of missingness in the post-treatment assessment. The test was not significant both in the total sample ($\chi2(15)=13.78$, $p=0.54$) and within each group (GC-MRT: $\chi2(4)=2.46$, $p=0.65$; SSRI: $\chi2(4)=6.85$, $p=0.14$; Control: $\chi2(15)=8.63$, $p=0.90$), strengthening the assumption that the data was missing completely at random.

TABLE S1. Completers vs. non-completers analysis. (a) full sample (n=105); (b) SSRI group (n=35).

(a)	Treatment Completers (n=90)		Treatment Non-completers (n=15)			
	Mean	SD	Mean	SD	t value	p value
Age (years)	29.90	7.24	35.60	11.92	1.80	0.09
LSAS Baseline	79.87	15.58	76.93	16.42	0.67	0.50
SPIN Baseline	48.53	8.57	48.20	8.24	0.14	0.89
PHQ Baseline	13.87	5.40	14.67	6.07	0.52	0.60

(b)	Treatment Completers (n=25)		Treatment Non-completers (n=10)			
	Mean	SD	Mean	SD	t value	p value
Age (years)	28.60	7.13	36.40	13.13	1.78	0.10
LSAS Baseline	82.00	15.29	75.10	11.44	1.29	0.21
SPIN Baseline	48.44	7.30	47.10	8.20	0.47	0.64
PHQ Baseline	13.84	5.74	14.20	6.73	0.16	0.87

FIGURE S1. Scatterplots of changes in dwell time on threat, by changes in social anxiety scores: (a) mid-treatment LSAS; (b) post-treatment LSAS; (c) mid-treatment SPIN; (d) post-treatment SPIN