RESEARCH ARTICLE

ANXIETY AND DEPRESSION ASSOCIATION OF AMERICA   WILEY

A randomized controlled trial of gaze-contingent music reward therapy for major depressive disorder

Dana Shamai-Leshem[1,2] ⬤ | Amit Lazarov[2] ⬤ | Daniel S. Pine[3] | Yair Bar-Haim[1,2] ⬤

[1]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

[2]School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel

[3]Section on Developmental Affective Neuroscience, National Institute of Mental Health, Bethesda, Maryland

Correspondence
Dana Shamai-Leshem, Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 69978, Israel.
Email: danasham@mail.tau.ac.il

Abstract

Background: Heightened attention allocation toward negative-valanced information and reduced attention allocation toward positive-valanced information represent viable targets for attention bias modification in major depressive disorder. Accordingly, we conducted a randomized controlled trial testing the efficacy of a novel gaze-contingent attention bias modification procedure for major depressive disorder.

Method: Sixty patients with major depressive disorder were randomly assigned to either eight training sessions of feedback-based gaze-contingent music reward therapy designed to divert patients' gaze toward positive over sad stimuli, or to a control condition which entailed eight sessions of gaze-noncontingent music. Clinician-rated and self-reported measures of depression, and proportion of dwell-time on sad faces, were assessed pretreatment, posttreatment, and at a 3-month follow-up.

Results: Gaze-contingent music reward therapy produced a greater reduction in dwell-time on sad faces compared with the control condition, but it failed to generalize to novel faces. Both groups manifested similarly significant reductions in depression symptoms from pre- to posttreatment that were maintained at follow-up. Exploratory analyses suggest that first-episode patients may benefit more from this therapy than patients with a history of multiple episodes.

Conclusions: Gaze-contingent music reward therapy can modify attention biases in depression, but clear differential clinical effects did not emerge. Theoretical and practical implications are discussed.

KEYWORDS
attention, attention allocation, attention bias, attention bias modification, depression, eye tracking, major depressive disorder

1 | INTRODUCTION

Models of major depressive disorder (MDD) suggest that attention biases contribute to the onset, maintenance, and recurrence of depressive episodes through engagement in elaborative processing of negative content (Armstrong & Olatunji, 2012; Beck, 1967, 1976, 2008; Farb, Irving, Anderson, & Segal, 2015). Corresponding research indicates that depressed individuals demonstrate an attentional bias toward negative stimuli (Armstrong & Olatunji, 2012; Duque & Vázquez, 2015; Gotlib & Joormann, 2010; Peckham, McHugh, & Otto, 2010), as well as a lack of a positive bias. That is, while nondepressed individuals attend more to positive stimuli, patients with MDD tend to divide their attention equally between positive and negative stimuli (Duque & Vázquez, 2015; Lazarov, Ben-Zion, Shamai, Pine, & Bar-Haim, 2018). Such biases have also been found in previously depressed individuals

       wileyonlinelibrary.com/journal/da   *Depress Anxiety.* 2021;38:134–145.

(Joormann & Gotlib, 2007; Newman, Quigley, Fernandez, Dobson, & Sears, 2019; Soltani et al., 2015) and may represent a risk factor for recurrence of future depressive episodes. These observations suggest viable targets for attention bias modification (ABM) therapy in depression.

ABM is a therapeutic approach designed to modify attention biases using computerized training (Bar-Haim, 2010; Beard, Sawyer, & Hofmann, 2012; Hakamata et al., 2010; Linetzky, Pergamin-Hight, Pine, & Bar-Haim, 2015). Small-to-medium effect sizes of ABM have been reported for anxiety disorders (Hakamata et al., 2010; Linetzky et al., 2015), whereas clinical efficacy in depression appears less consistent, possibly reflecting limitations in reaction-time (RT)-based training (Armstrong & Olatunji, 2012).

A recently developed ABM protocol, gaze-contingent music-reward therapy (GC-MRT), addresses some of the limitations of RT-based training by applying eye-tracking technology. In GC-MRT, patients view matrices of negative and neutral faces while listening to music of their choice, which is played only when patients' gaze is fixated on neutral faces. GC-MRT reduces attention bias and social anxiety (Lazarov, Pine, & Bar-Haim, 2017; Linetzky, Kahn, Lazarov, Pine, & Bar-Haim, 2019). Because depressed patients dwell longer on sad faces and lack a positive bias toward happy faces (Lazarov et al., 2018), we applied a modified GC-MRT to patients with MDD. The present randomized controlled trial (RCT) examines the clinical efficacy and associated mechanism of GC-MRT for patients with MDD, targeting enhanced dwelling on sad faces and reduced dwelling on happy faces. Patients were randomly assigned to either GC-MRT or a control condition entailing music noncontingent on viewing patterns.

We expected that relative to patients in the control group, patients receiving GC-MRT would demonstrate greater reductions in dwell-time on sad faces and MDD symptoms that would be sustained at a 3-month follow-up. Of note, multiple past depressive episodes have been shown to predict poor treatment response (Gorwood et al., 2010) and the presence of various cognitive deficits (e.g. Basso & Bornstein, 1999; Elgamal, Denburg, Marriott, & Macqueen, 2010; Vanderhasselt & De Raedt, 2009). Hence, we also expected that group differences in symptom reduction would be greater among patients experiencing their first depressive episode relative to patients who had experienced prior episodes. This latter hypothesis was tested in an exploratory manner, due to the sample size and the absence of prior results with GC-MRT.

2 | METHODS

2.1 | Participants

CONSORT diagram appears in Figure 1. Participants were 60 patients with MDD (Mage = 41.84 years, range = 18–65, 26 females). Inclusion criteria were: (a) MDD diagnosis; (b) total score ≥ 18 on the Montgomery–Asberg Depression Rating Scale (MADRS); (c) 18–65 years of age; and (d) normal or corrected-to-normal vision. Exclusion

criteria: (a) history or current psychosis, bipolar disorder, manic or hypomanic episode; (b) epilepsy or brain injury; (c) severe suicidal ideation; (d) drugs or alcohol abuse; (e) pharmacological treatment not stabilized for at least 3 months or concurrent psychotherapy; and (f) eye-tracking calibration difficulties. Participants were randomly assigned to either GC-MRT (n = 30, 13 females) or to a control condition (n = 30, 13 females). Some participants had comorbidities: 6 panic disorder (2 in GC-MRT), 1 agoraphobia (GC-MRT), 23 social anxiety disorder (8 in GC-MRT), 2 obsessive-compulsive disorder (1 in GC-MRT), and 32 generalized anxiety disorder (16 in GC-MRT). Among all participants, 18 (10 in GC-MRT) were in the midst of their first depressive episode. The study was approved by the Tel Aviv University Ethics Committee. Written informed consent was provided by all participants (ClinicalTrials.gov identifier: NCT02945735).

2.2 | Diagnostic and self-report measures

The Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001) was used for initial phone-screening. Those with scores ≥10, indicating at least moderate depression (Kroenke et al., 2001) were invited to an in-person clinical interview. The PHQ-9 has good psychometric properties (Kroenke et al., 2001).

Mini-International Neuropsychiatric Interview (M.I.N.I; Sheehan et al.,1998), was used to determine MDD and comorbid diagnoses. It is a structured interview for DSM-IV and ICD-10 diagnoses, with good reliability, sensitivity, and specificity (Lecrubier et al., 1997; Sheehan et al., 1997).

Montgomery–Åsberg Depression Rating Scale (MADRS; Montgomery & Asberg, 1979) is a 10-item clinician administered scale used to diagnose depression according to DSM-IV. Higher scores indicate greater depression. A cut-off score of 18—moderate depression (Müller, Szegedi, Wetzel, & Benkert, 2000), was used as an inclusion criterion. MADRS total score was the study's primary outcome. Mean Cronbach's α in the current sample was .73.

The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996), is a self-reported 21-item inventory measuring MDD severity. The BDI-II has good reliability and internal consistency (Beck et al., 1996; Sprinkle et al., 2002), and served as the secondary outcome. Mean Cronbach's α in our sample was .84.

The Credibility/Expectancy Questionnaire (CEQ; Devilly & Borkovec, 2000), is a 6-item scale that measures expectancy of clinical improvement and perceived treatment credibility. Items are ranked on a 9-point scale or 0–100%. In both cases, higher scores indicate higher expectancy/credibility. The CEQ has high internal consistency and good test–retest reliability (Devilly & Borkovec, 2000). Cronbach's α in the current sample was .83. The CEQ was administered at pretreatment, after a comprehensive explanation of the study's rational and procedures, and before randomization. Although the participants were fully informed that there are active and sham conditions in the study and that they will be randomly assigned to one of these, to keep patients blind to their specific group assignment the exact nature of the differences between the two
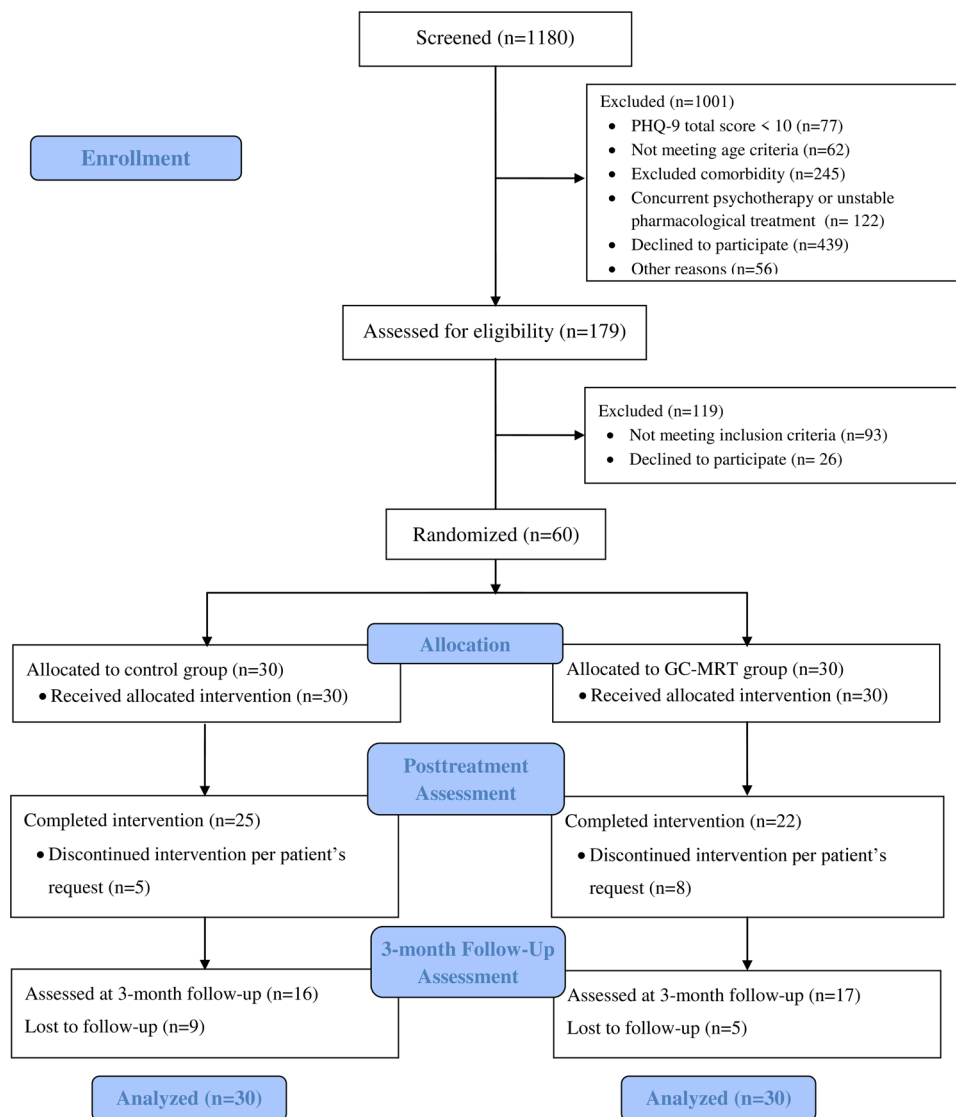
FIGURE 1 CONSORT diagram of participants' progress through the study. GC-MRT, gaze-contingent music reward therapy; PHQ-9, Patient Health Questionnaire

conditions (i.e., the exact contingency between eye-gaze and music) was kept concealed. Thus, the CEQ scores we report reflect pretreatment expectancies before it began.

2.3 | Attention allocation measurement

Attention allocation was assessed using a free-viewing eye-tracking task (Lazarov, Abend, & Bar-Haim, 2016; Lazarov et al., 2018). Each trial began with a fixation cross until a 1,000 ms fixation was recorded. Then, a matrix of 16 faces, half sad and half happy (Figure 2), appeared for 6,000 ms followed by an intertrial interval of 2,000 ms. The task consisted of two blocks of 30 trials each, with a different matrix appearing in each trial. Participants were instructed to look at the matrices in any way they like. For a complete description of the measurement task see Lazarov et al. (2018). This assessment task demonstrated high internal

consistency (Cronbach's αs ranging .88-.96) and good test–retest reliability ($rs = .72$–0.74) in depressed participants (Lazarov et al., 2018). Mean Cronbach's α for percent dwell-time on sad faces in the current sample was .91.

2.4 | Treatment conditions

2.4.1 | GC-MRT

Participants in this condition received eight 20-min sessions twice a week, designed to divert attention away from sad stimuli and toward happy stimuli (see Lazarov et al., 2017). At the beginning of each session, participants selected a 12-min music track they wanted to listen to during the session. Thirty different matrices were presented for 24 s each, with no intertrial intervals. Patients heard their selected music only when fixating on the happy faces.

FIGURE 2 A single matrix, comprised of eight sad and eight happy faces

2.4.2 | Control condition

Participants underwent the same procedure as in GC-MRT with the music of their choice played without interruptions, that is, not contingent upon their gaze.

2.5 | Apparatus and eye-tracking measures

Eye-tracking data were recorded using RED500 and analyzed using BeGaze software (SensoMotoric Instruments, Inc., Teltow, Germany). Stimuli were presented on a 22-inch monitor (screen resolution 1680×1050). Operating distance to monitor was 70 cm. Sampling rate was 500 Hz.

Following Lazarov et al. (2017, 2018), two areas of interest (AOIs) were defined for each matrix: one consisting of the eight sad faces (sad AOI) and one of the eight happy faces (happy AOI). Total dwell-time per AOI was calculated as the total dwell-time averaged across all matrices. Percent dwell-time on the sad AOI was calculated as the proportion of the averaged dwell-time on the sad AOI relative to dwell-time on both happy and sad AOIs. A score above 50%

reflects longer dwelling on the sad AOI whereas a score below 50% reflects longer dwelling on the happy AOI.

2.6 | Procedure

Study design was a double-blind RCT with two groups (GC-MRT, Control), such that both the independent evaluators and participants were blind to group allocation. Potential participants were phone screened for MDD symptoms using the PHQ-9 (Kroenke et al., 2001). Those with a score ≥10 were invited to an in-person clinical interview. Candidates meeting inclusion criteria were invited to a pre-treatment assessment in which attention bias was measured using the free-viewing task. Then, participants were randomly assigned to eight sessions of either GC-MRT or control, both delivered twice a week. Generalization of learning was tested using a different set of faces than used for training.

One week following the last therapy/control session, a post-treatment assessment was held. Attention bias was measured again, and clinical status was re-assessed. In addition, patients in the GC-MRT group were asked whether they had explicit knowledge of the

training rule. Three months later, the same assessments were repeated to test longer-term treatment effects. Participants in the control condition were given the opportunity to receive GC-MRT after the study ended. Data collection was carried out between November 2016 and January 2019.

2.7 | Data analysis

Independent samples *t* tests compared groups' descriptive statistics at pretreatment. Clinical effects were analyzed using the intent-to-treat principle, by deploying random effect time-series models in generalized estimating equations (GEE; Zeger & Liang, 1986; Zeger, Liang, & Albert, 1988). GEE considers correlations between repeated measurements while addressing missing data via estimated marginal means relying on the entire sample (all randomized participants). Wald's chi-square (Rotnitzky & Jewell, 1990) was used to test whether the coefficient of the predictors in the models were significantly different from zero. Clinician-rated (MADRS) and self-reported (BDI-II) effects on depression were estimated using models containing the main effects of time (pretreatment, posttreatment, follow-up), group (GC-MRT, Control) and their interactions. Secondary GEE analyses applied to patients who completed the study and had full data for either pre- and posttreatment or posttreatment and follow-up are described in the Supplemental Material. Finally, the effect of depressive episodes history (i.e., first-episode vs. multiple episodes) was examined using GEE analyses as above, introducing depressive history as a predictor.

To compare attention allocation patterns as a function of treatment condition, percent dwell-time on sad faces was estimated using a GEE model containing the main effects of time (eight training sessions), treatment condition (GC-MRT, control), and their interaction. Near-transfer of training to novel faces was estimated using percent dwell-time on novel sad faces in a GEE model considering the main effects of time (pretreatment, posttreatment, and follow-up), group (GC-MRT, control), and their interaction. See the Supplemental Material for results of additional models examining dwell-time on sad and happy faces separately, using face type (happy/sad) as another predictor. GEE analyses of percent dwell-time on sad faces during the training sessions and in the three assessment points for completers data are also provided in Supplemental Material.

To examine whether individual differences in near-transfer to novel faces were correlated with clinical improvement, Pearson's correlations were computed between change in clinical outcomes (MADRS and BDI-II) and change in percent dwell-time on sad faces (subtraction scores) from pre- to posttreatment and from pretreatment to follow-up.

To examine possible effects of expectancy and treatment credibility on clinical improvement, Pearson's correlations were computed between pretreatment expectancy and credibility scores and pre- to posttreatment change in clinical outcomes (MADRS and BDI-II). Finally, to evaluate possible mediators of treatment outcomes, additional exploratory analyses were conducted using gender (male and female), comorbid generalized anxiety (diagnosed and not diagnosed) and comorbid social anxiety (diagnosed and not diagnosed) as potential predictors, and age as a covariate in the above-described models. All statistical tests were two-sided, with $\alpha \leq .05$.

3 | RESULTS

3.1 | Preliminary analyses

The groups did not differ on age, gender distribution, education, first versus multiple depressive episodes, clinician or self-rated depression, and percent dwell-time on sad faces at pretreatment (all *p*s > .17, Table 1). In addition, groups did not differ in the number of participants with current (GC-MRT: *n* = 3, control: *n* = 8) or past (GC-MRT: *n* = 21, control: *n* = 22) pharmacological treatment (*p*s > .09).

3.2 | Clinical change in major depression symptoms following treatment

3.2.1 | Primary outcome

Figure 3a depicts the results of the GEE model for MADRS scores. Analysis revealed a main effect of time (Wald $\chi^2(2) = 46.79$, *p* < .0001), as well as nonsignificant effects of group (Wald

TABLE 1 Demographic characteristics, depression and anxiety symptoms, and attention bias scores by group at pretreatment, posttreatment, and follow-up

	GC-MRT group		Control group	
Variable	M	SD	M	SD
Age (years)	43.37	10.89	40.33	12.91
Years of education	14.3	2.76	13.67	2.4
MADRS score at pretreatment	28.6	4.43	30.4	5.14
MADRS score at posttreatment	24.1	7.03	24.53	8.4
MADRS score at follow-up	20.36	9.69	20.25	10.95
BDI-II score at pretreatment	29.94	9.66	30.76	8.71
BDI-II score at posttreatment	24.096	8.82	22.68	10.66
BDI-II score at follow-up	18.67	13.18	20.14	11.92
Dwell-time on sad faces at pretreatment (%)	49.56	3.07	49.35	2.64
Dwell-time on sad faces at posttreatment (%)	43.66	11.46	45.63	7.25
Dwell-time on sad faces at follow-up (%)	45.13	10.97	46.32	7.62

Abbreviations: BDI-II, Beck Depression Inventory-II; GC-MRT, gaze-contingent music reward therapy; MADRS, Montgomery–Asberg Depression Rating Scale.
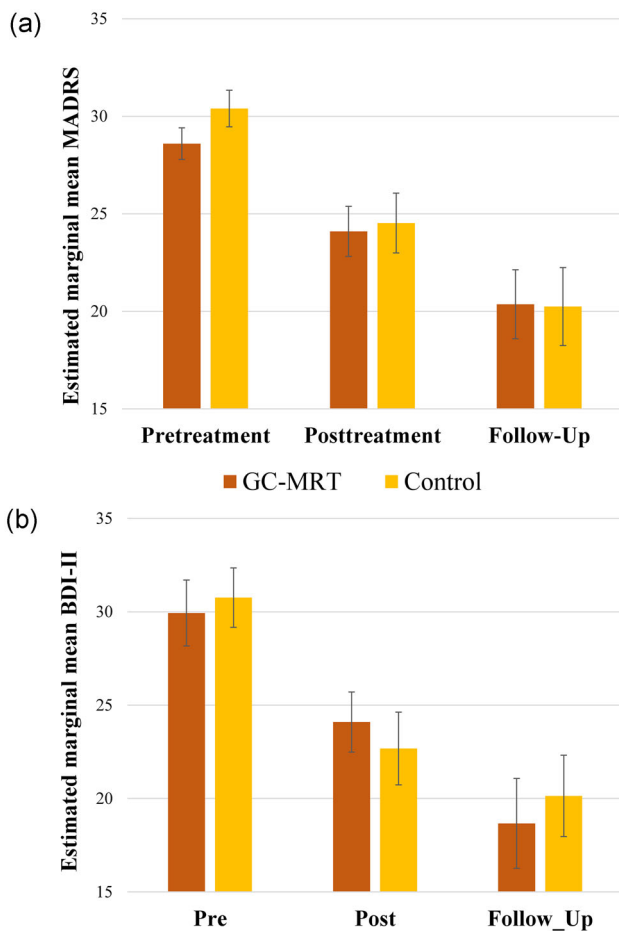
FIGURE 3 MADRS (a) and BDI-II (b) scores by group and time. Higher values indicate greater depression. Error bars denote the standard error of the estimated mean. BDI-II, Beck Depression Inventory-II; GC-MRT, gaze-contingent music reward therapy; MADRS, Montgomery–Asberg Depression Rating Scale

$\chi^2(1) = 0.23$, $p = .629$), and group × time (Wald $\chi^2(2) = 0.59$, $p = .743$). Follow-up analysis revealed a reduction of symptoms from pre- to posttreatment ($p < .0001$, $d = 0.80$), and from posttreatment to follow-up ($p < .01$, $d = 0.44$).

3.2.2 | Secondary outcome

Figure 3b depicts the results of the GEE model for the BDI-II scores. Analysis indicated a main effect of time (Wald $\chi^2(2) = 44.10$, $p < .00001$), as well as a nonsignificant effects of group (Wald $\chi^2(1) = 0.02$, $p = .887$), and group × time (Wald $\chi^2(2) = 1.20$, $p = .548$). Follow-up analysis revealed a reduction from pre- to posttreatment ($p < .0001$, $d = 0.73$) and from posttreatment to follow-up ($p < .01$, $d = 0.35$).

3.2.3 | First versus multiple depressive episodes

Table 2 presents the results for GEE models (MADRS and BDI-II) for first versus multiple episodes. For the MADRS scores, a nonsignificant trend-level group × time × depression history interaction emerged (Wald $\chi^2(2) = 5.22$, $p = .073$). Follow-up analysis revealed a trend-level time × depression history interaction effect in the GC-MRT group (Wald $\chi^2(2) = 5.95$, $p = .051$, Figure 4a), but not in the control group (Wald $\chi^2(2) = 1.85$, $p = .397$, Figure 4c). In the GC-MRT group, MADRS depression scores decreased from pre- to posttreatment in first-episode patients ($p < .001$, $d = 1.57$) but not in patients with multiple past episodes ($p = .208$). However, the difference between first versus multiple episode patients at posttreatment was at a nonsignificant trend level ($p = 0.061$). MADRS scores remained stable from posttreatment to follow-up in first-episode patients ($p = .335$), whereas patients with multiple past episodes exhibited an additional reduction in symptoms ($p = .039$, $d = 0.52$), resulting in a nonsignificant difference between patients with first versus multiple episodes at follow-up ($p = .34$). Among first-episode patients, MADRS depression scores did not differ between GC-MRT and control at posttreatment ($p = .181$), but did differ at follow-up ($p = .029$, $d = 0.997$).

The same analysis for BDI-II scores revealed a significant effect of time (Wald $\chi^2(2) = 61.65$, $p < .0001$), that was qualified by a group × time × depression history interaction effect (Wald $\chi^2(2) = 11.57$, $p < .01$). Follow-up analysis revealed a time × depression history interaction in the GC-MRT group (Wald

TABLE 2 Depression estimated scores by group at pretreatment, posttreatment, and follow-up among patients with first versus multiple depressive episodes

| | GC-MRT group | | | | | | Control group | | | | | |
| | Pre | | Post | | Follow-up | | Pre | | Post | | Follow-up | |
Variable	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
MADRS												
F. Episode	29.30	2.33	21.27	6.83	18.12	9.46	31.63	6.69	27.26	11.12	25.58	4.72
M. Episodes	28.25	5.14	25.62	6.32	21.42	9.61	29.95	4.35	23.73	7.33	18.50	11.46
BDI-II												
F. Episode	32.71	9.51	18.77	7.91	15.05	7.56	30.47	10.86	25.08	11.05	23.85	8.57
M. Episodes	28.55	9.44	26.71	8.11	20.34	14.00	30.87	7.79	21.94	10.07	18.86	12.08

Abbreviations: BDI-II, Beck Depression Inventory-II; F. Episode, first-episode; GC-MRT, gaze-contingent music reward therapy; M. Episode, multiple episodes; MADRS, Montgomery–Asberg Depression Rating Scale.
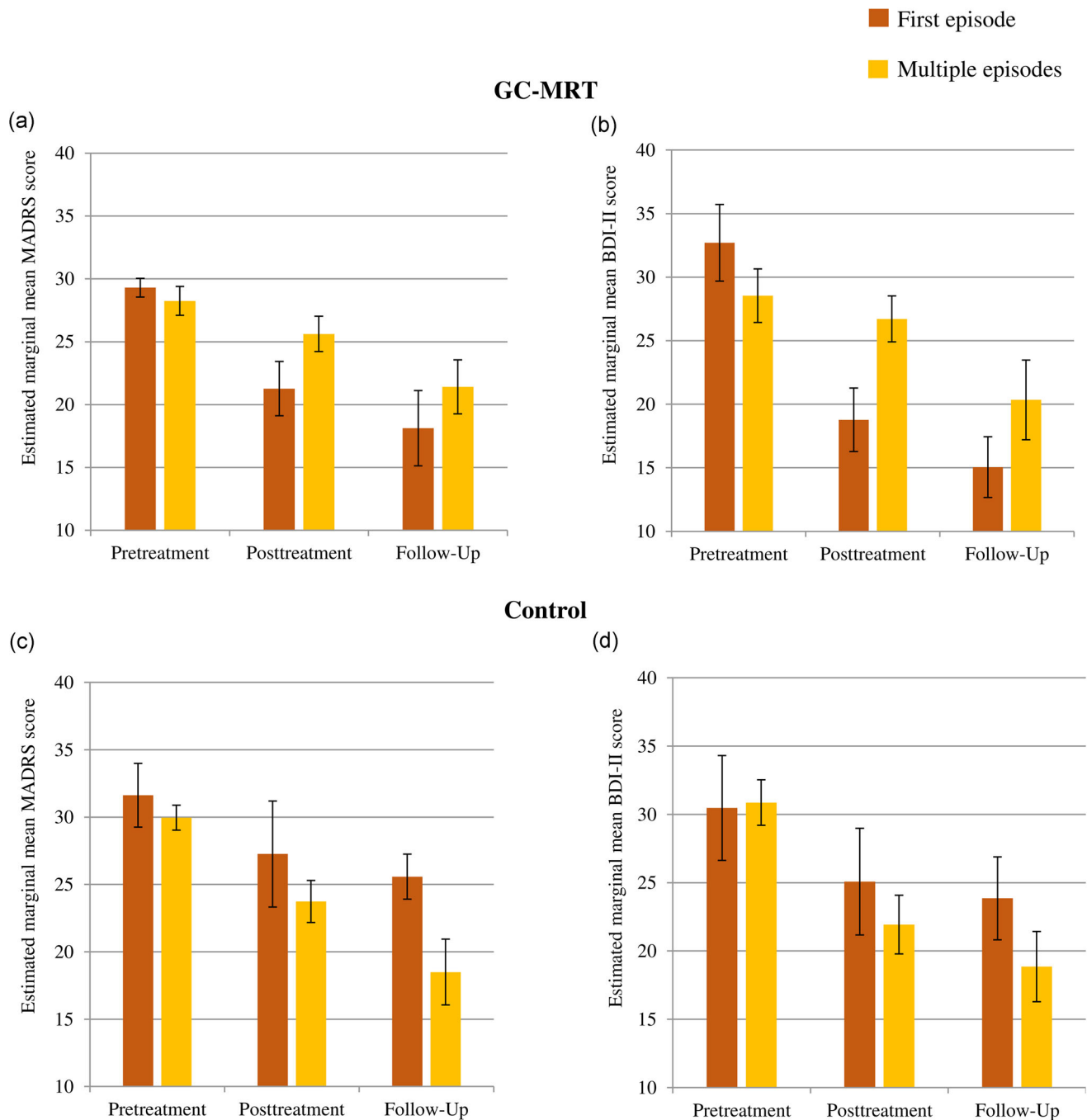
FIGURE 4 MADRS and BDI-II scores in the GC-MRT (a and b) and control (c and d) groups by time by history of depressive episodes. Higher values indicate greater depression. Error bars denote the standard error of the estimated mean. BDI-II, Beck Depression Inventory-II; GC-MRT, gaze-contingent music reward therapy; MADRS, Montgomery–Asberg Depression Rating Scale

$\chi^2(2) = 9.61$, $p < .01$, Figure 4b), but not in the control group (Wald $\chi^2(2) = 1.91$, $p = .385$, Figure 4d). In the GC-MRT group, BDI-II scores decreased from pre- to posttreatment in first-episode patients ($p < .0001$, $d = 1.59$) but not in patients with multiple past episodes ($p = .324$), and a significant difference emerged between patients with first versus multiple episodes at posttreatment ($p = .012$, $d = 0.98$). Finally, while BDI-II scores remained stable from posttreatment to follow-up in first-episode patients ($p = .173$), patients with multiple past episodes exhibited a reduction in symptoms ($p = .022$, $d = 0.56$). A nonsignificant difference emerged between patients with first versus multiple episodes at follow-up ($p = .291$). Among first-episode patients, BDI-II depression scores did not differ between GC-MRT and control at posttreatment ($p = .174$), but did differ at follow-up ($p = .023$, $d = 1.09$).

3.3 | Change in attention allocation across treatment sessions

Average dwell-time on sad faces in the first five matrices of the first treatment session was used to compare baseline performance of the GC-MRT ($M = 44.98$, $SD = 10.27$) and control ($M = 46.66$, $SD = 9.32$) groups, which did not differ at baseline ($t(56) = −0.59$, $p = .56$). Results of the GEE model for percent dwell-time on sad faces across training sessions are depicted in Figure 5a. Main effects of group (Wald $\chi^2(1) = 13.61$, $p < .001$), and session (Wald $\chi^2(8) = 35.32$, $p < .0001$), were qualified by a group × session interaction (Wald $\chi^2(8) = 42.13$, $p < .00001$), reflecting differential learning in the two groups. Follow-up analysis indicated that while no difference between groups in percent dwell-time on sad faces was evident in the first training session ($t(57) = −0.44$, $p = .66$), a significant between-groups difference was manifested in Sessions 2–8 (all $ps < .01$). Percent dwell-time on sad faces decreased by 17.74% between Sessions 1–8 in the GC-MRT group ($p < .0001$, $d = 0.91$), whereas no learning was evident in the control group ($p = .14$). A significant linear and quadratic trends emerged in the GC-MRT group ($Fs(1,20) = 7.74$ and 5.44, $ps < .01$ and .03, respectively), but not in the control group (both $ps > .49$).

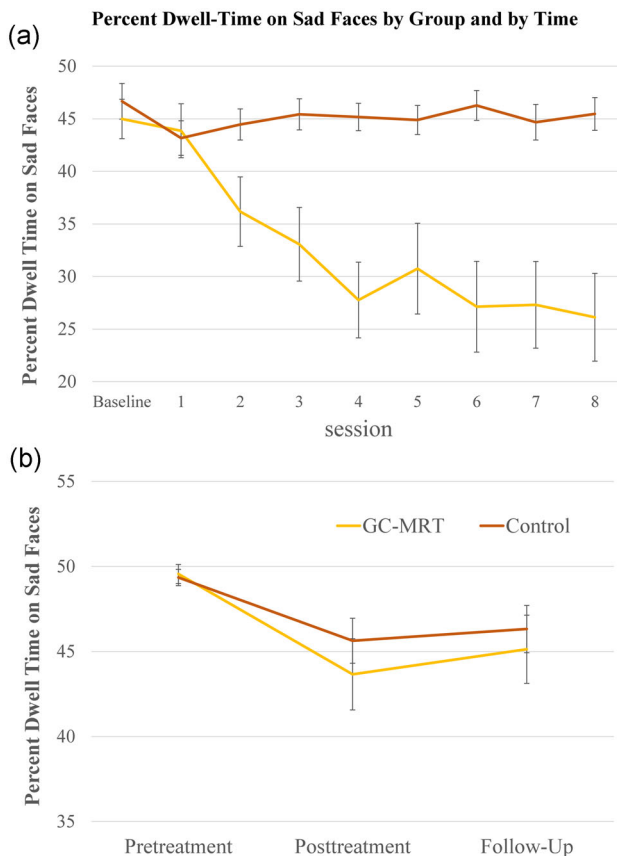(a) **Percent Dwell-Time on Sad Faces by Group and by Time**

(b)

FIGURE 5 Percent dwell-time on sad faces by group and treatment session (a; Session 1–8) and percent dwell-time on sad faces not used in training by group and assessment session (b; pretreatment, posttreatment, and follow-up). GC-MRT, gaze-contingent music reward therapy

Additional analyses examining dwell-time (ms) for each face type separately revealed that GC-MRT training led to both decreased dwell-time on sad faces and increased dwell-time on happy faces (see Supplemental Material for complete analyses and results).

3.4 | Change in attention allocation to novel faces

Figure 5b presents the results of the GEE model for percent dwell-time on sad faces not used in training at pretreatment, posttreatment, and follow-up. Analysis revealed a main effect of time (Wald $\chi^2(2) = 14.94$, $p < .01$), and nonsignificant effects of group (Wald $\chi^2(1) = 0.43$, $p = .51$), and group × time interaction (Wald $\chi^2(2) = 0.74$, $p = .69$). Follow-up analysis revealed reduction in percent dwell-time on sad faces from pre- to posttreatment ($p < .001$, $d = 0.68$), and no change from posttreatment to follow-up ($p = .31$, $d = −0.11$).

Pearson's correlations between changes in percent dwell-time on sad faces and MDD symptoms from pre- to posttreatment and from pretreatment to follow-up revealed nonsignificant correlations for the MADRS ($rs(48) = 0.18$ and $.32$, $ps = .22$ and $.08$, respectively), and significant correlations for the BDI-II ($rs(47) = 0.36$ and $.38$, $ps = .01$ and $.04$, respectively).

3.5 | GC-MRT rule-learning

Thirteen participants in the GC-MRT group reported that they had learned the contingency between face emotion and music embedded in the training task. GEE analysis of percent dwell-time on sad faces across the training sessions within the GC-MRT group, with explicit rule-learning (yes, no) as a between-subjects variable and Session (1–8) as a within-subjects variable, revealed a main effects of session (Wald $\chi^2(8) = 104.65$, $p < .0001$), and rule-learning (Wald $\chi^2(1) = 76.67$, $p < .0001$), which were subsumed under a rule-learning × session interaction effect (Wald $\chi^2(8) = 100.29$, $p < .0001$; see Figure S1). Follow-up analysis indicated that while no difference in percent dwell-time on sad faces between explicit rule learners and non-learners was evident in the first training session ($t(17.46) = 1.83$, $p = .08$), a significant between-groups difference was manifested in Sessions 2–8 (all $ps < .01$). Percent dwell-time on sad faces decreased by 26.7% from Session 1–8 among those who explicitly reported the rule ($p < .0001$, $d = 1.41$), whereas no learning was indicated among those who did not ($p = .77$). A significant linear and quadratic trends emerged among those who explicitly learned the contingency ($Fs(1,10) = 15.06$ and 41.50, $ps < .01$ and $.001$, respectively), but not among those who did not ($Fs(1,9) = 0.27$ and 4.80, $ps = .614$ and $.056$, respectively).

Importantly, however, even those who explicitly learned the contingency did not generalize this knowledge to new faces (Wald $\chi^2(2) = 2.63$, $p = .268$, Figure S1), and had no advantage over those who did not report the rule in clinician-rated (MADRS, Wald $\chi^2(2) = 1.81$, $p = .40$), or self-reported (BDI-II, Wald $\chi^2(2) = 4.23$, $p = .12$) depression.

3.6 | Effects of expectancy and treatment credibility on clinical outcome

Following explanation at baseline, patients found the treatment rational moderately credible ($M = 6.07$, $SD = 1.90$) and expected a mean of ~54% improvement in their symptoms. Importantly, expectancy and credibility were not associated with treatment outcomes (MADRS or BDI-II), rs range = −.29 to .20, all $ps > .10$.

3.7 | Additional potential mediators

Models that included gender, comorbid generalized anxiety, and comorbid social anxiety as additional predictors yielded no difference from the above-detailed results for MADRS, BDI-II, and percent dwell-time on sad faces during assessment sessions (all $ps > .20$). Models that included age as a covariate yielded no difference from the original results for MADRS and percent dwell-time on sad faces during assessment sessions (all $ps > .08$). For self-reported BDI-II scores, a significant group × time × age interaction effect emerged (Wald $\chi^2(2) = 8.07$, $p = .018$). However, follow-up analyses, introducing age as a covariate, revealed no differences between groups at posttreatment or at follow-up ($Fs = 0.20$ and 0.39, $ps = .66$ and $.54$, respectively).

4 | DISCUSSION

This study examined the efficacy of GC-MRT, a novel gaze-contingent feedback-based therapy, for patients with MDD. Results indicate effective target engagement in the GC-MRT group, reflected in reduced dwelling on sad faces over time, but no near-transfer generalization of learning to new faces. Reduction in depressive symptoms occurred in both groups with no advantage for GC-MRT. While some evidence emerged for moderation by past depression history, results only emerged in a posthoc analysis and mainly for the secondary outcome measure.

Lack of group differences in symptom reduction may be attributed to lack of learning generalization in the GC-MRT group. Generalization of training is essential for clinical efficacy in ABM. Modification of behavior during training sessions alone is unlikely to result in far-transfer therapeutic effect (Hertel & Mathews, 2011). Indeed, in the current study, change in bias toward novel sad faces was correlated with self-reported depression severity, suggesting that individuals who better generalized the learning contingency also manifested greater far-transfer therapeutic effect. Previous GC-MRT studies in anxiety patients did show group differences in near-transfer effects and in symptom reduction (Lazarov et al., 2017; Linetzky et al., 2019). In contrast, the depressed patients in the current study showed no generalization of learning to untrained faces. This difference in findings may be related to deficient reward processing in MDD (Eshel & Roiser, 2010; Nestler & Carlezon, 2006; Whitton, Treadway, & Pizzagalli, 2015). More specifically, Anderson,

Leal, Hall, Yassa, and Yantis (2014) demonstrated that while depressed individuals were capable of learning associations between stimuli and reward, they later failed to modulate their attention to such associations outside the training sessions. Similarly, in the present study, even participants in the GC-MRT group that managed to learn the reward contingency during training failed to generalize the learned contingency to novel faces outside the treatment sessions. To shed light on the mechanisms underlying such learning deficits, future studies could compare the GC-MRT-related learning and generalization processes of patients with MDD and nondepressed individuals.

While in the current study symptoms reduction did not differ between groups, a significant symptom reduction was evident in both, which could have occurred for many reasons. First, this may reflect spontaneous remission characterizing the natural course of depressive episodes (Whiteford et al., 2013). However, this account is somewhat unlikely given that the observed symptom reductions occurred after a 4-week treatment, whereas spontaneous remission in MDD typically occurs after longer time periods (Posternak et al., 2006). Moreover, the persistence of improvement at posttreatment follow-up also suggests a change in symptom trajectories related to some aspect of study participation. Second, it is possible that similar expectations for improvement in both groups drove clinical results. Pretreatment expectancies have been found to predict symptoms change in MDD (Webb, Kertz, Bigda-Peyton, & Björgvinsson, 2013), and evidence points to strong placebo response among depressed patients (Rief et al., 2009). Importantly, however, in the current study pretreatment expectancies were not associated with symptoms change, therefore somewhat diminishing the likelihood of this explanation. Third, a common therapeutic mechanism could conceivably be shared across the two treatment conditions. One possible mechanism is the therapeutic effect of music per se (Erkkilä et al., 2011; Maratos, Gold, Wang, & Crawford, 2008), which has been shown to improve mood and increase arousal (Juslin & Sloboda, 2013; Salimpoor, Benovoy, Longo, Cooperstock, & Zatorre, 2009). Moreover, assuming that music itself indeed has a therapeutic effect, this could explain the lack of group difference in clinical outcomes. Specifically, while participants in the control group listened to music continuously without interruptions, participants in the GC-MRT group listened to interrupted music based on their gaze patterns. Thus, the therapeutic effect of GC-MRT might have been masked by an enhanced effect of music in the control condition. Future studies could attempt to tackle this possibility by using different control conditions better controlling for music exposure (e.g., yoked music between participants in the different groups or a no music control condition).

The results do not indicate superior clinical efficacy for GC-MRT over a control condition. Nevertheless, additional analyses suggest that GC-MRT may have enhanced efficacy for patients dealing with their first depressive episode, as indicated by self-reported depression (not observed when using the clinician-rated measure). This tentative finding is in line with Gorwood et al. (2010), who found that multiple depressive episodes are associated with poor treatment

response in MDD. GC-MRT requires cognitive control and relies on reward processing, both of which have been associated with the number of previous depressive episodes (Morgan et al., 2016; Vanderhasselt & De Raedt, 2009). Given the exploratory nature of these analyses in the current study, future research should further test whether GC-MRT may be more efficacious for patients on their first depressive episode and devise training protocols that improve near transfer of training to other faces.

The current study carries some limitations. First, since previous research has found both attention bias toward sad faces and away from happy faces (Duque & Vázquez, 2015; Lazarov et al., 2018), our training and measurements included both facial expressions presented concurrently. However, a study on ABM for patients with MDD demonstrated a near-transfer effect for sad-neutral faces, but not for happy-neutral faces (Beevers, Clasen, Enock, & Schnyer, 2015). It is possible that by presenting sad and happy facial expressions simultaneously this direct effect was blunted. Future studies could use matrices of neutral and emotional faces (e.g., sad-neutral) to maximize independent effects. Second, the current study has a modest sample size that may possess limited power to detect small between-group differences. Relatedly, the current study is potentially underpowered to systematically explore the effect of depression history on treatment outcome. We preliminarily examined the effect of first versus multiple depressive episodes on treatment outcome with no direct control over the number of first-episode patients in the study. Future research is advised to control for this important variable and ascertain larger numbers of participants with different depression histories.

5 | CONCLUSION

This RCT is the first to examine efficacy of GC-MRT for MDD. No specific treatment effects were found, though preliminary findings tentatively suggest a potential moderation by depression history. Lack of near-transfer learning may account for the failure to observe group differences, reflecting deficits in reward learning among MDD patients. Future studies might adapt the current GC-MRT procedures in an attempt to improve learning generalization among MDD patients or to bypass this potential learning-related deficit. As compared with anxiety disorders, more research on underlying mechanism is needed to inform refinements in GC-MRT for MDD.

CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT
The deidentified data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

ORCID
Dana Shamai-Leshem http://orcid.org/0000-0003-1397-9214
Amit Lazarov http://orcid.org/0000-0002-2811-2739
Yair Bar-Haim https://orcid.org/0000-0002-4630-9180

REFERENCES
Anderson, B. A., Leal, S. L., Hall, M. G., Yassa, M. A., & Yantis, S. (2014). The attribution of value-based attentional priority in individuals with depressive symptoms. *Cognitive, Affective, & Behavioral Neuroscience*, 14(4), 1221–1227. https://doi.org/10.3758/s13415-014-0301-z

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, 32(8), 704–723. https://doi.org/10.1016/j.cpr.2012.09.004

Bar-Haim, Y. (2010). Research review: Attention bias modification (ABM): A novel treatment for anxiety disorders. *Journal of Child Psychology and Psychiatry*, 51(8), 859–870. https://doi.org/10.1111/j.1469-7610.2010.02251.x

Basso, M. R., & Bornstein, R. A. (1999). Relative memory deficits in recurrent versus first-episode major depression on a word-list learning task. *Neuropsychology*, 13(4), 557–563. https://doi.org/10.1037/0894-4105.13.4.557

Beard, C., Sawyer, A. T., & Hofmann, S. G. (2012). Efficacy of attention bias modification using threat and appetitive stimuli: A meta-analytic review. *Behavior Therapy*, 43(4), 724–740. https://doi.org/10.1016/J.BETH.2012.01.002

Beck, A. T. (1967). *Depression: Causes and treatment*. Philadelphia, PA: University of Pennsylvania Press.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. New York: International Universities Press.

Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry*, 165(8), 969–977. https://doi.org/10.1176/appi.ajp.2008.08050721

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Bdi-ii, beck depression inventory: Manual (Second). San Antonio. TX: Psychological Corporation.

Beevers, C. G., Clasen, P. C., Enock, P. M., & Schnyer, D. M. (2015). Attention bias modification for major depressive disorder: Effects on attention bias, resting state connectivity, and symptom change. *Journal of Abnormal Psychology*, 124(3), 463–475. https://doi.org/10.1037/abn0000049

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, 31(2), 73–86. https://doi.org/10.1016/S0005-7916(00)00012-4

Duque, A., & Vázquez, C. (2015). Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking study. *Journal of Behavior Therapy and Experimental Psychiatry*, 46, 107–114. https://doi.org/10.1016/j.jbtep.2014.09.005

Elgamal, S., Denburg, S., Marriott, M., & Macqueen, G. (2010). Clinical factors that predict cognitive function in patients with major depression. *The Canadian Journal of Psychiatry*, 55, 653–661.

Erkkilä, J., Punkanen, M., Fachner, J., Ala-Ruona, E., Pöntiö, I., Tervaniemi, M., ... Gold, C. (2011). Individual music therapy for

depression: Randomised controlled trial. *British Journal of Psychiatry*, *199*(2), 132–139. https://doi.org/10.1192/bjp.bp.110.085431

Eshel, N., & Roiser, J. P. (2010). Reward and punishment processing in depression. *Biological Psychiatry*, *68*(2), 118–124. https://doi.org/10.1016/J.BIOPSYCH.2010.01.027

Farb, N. A. S., Irving, J. A., Anderson, A. K., & Segal, Z. V. (2015). A two-factor model of relapse/recurrence vulnerability in unipolar depression. *Journal of Abnormal Psychology*, *124*(1), 38–53. https://doi.org/10.1037/abn0000031

Gorwood, P., Rouillon, F., Even, C., Falissard, B., Corruble, E., & Moran, P. (2010). Treatment response in major depression: Effects of personality dysfunction and prior depression. *British Journal of Psychiatry*, *196*(2), 139–142. https://doi.org/10.1192/bjp.bp.109.067058

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Review of Clinical Psychology*, *6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305.Cognition

Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J. C., Fox, N. A., Leibenluft, E., ... Pine, D. S. (2010). Attention bias modification treatment: A meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, *68*(11), 982–990. https://doi.org/10.1016/j.biopsych.2010.07.021

Hertel, P. T., & Mathews, A. (2011). Cognitive bias modification: Past perspectives, current findings, and future applications. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *6*(6), 521–536. https://doi.org/10.1177/1745691611421205

Joormann, J., & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of Abnormal Psychology*, *116*(1), 80–85. https://doi.org/10.1037/0021-843X.116.1.80

Juslin, P. N., & Sloboda, J. A. (2013). Music and emotion. *The Psychology of Music*, 583–645. https://doi.org/10.1016/B978-0-12-381460-9.00015-8

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lazarov, A., Abend, R., & Bar-Haim, Y. (2016). Social anxiety is related to increased dwell time on socially threatening faces. *Journal of Affective Disorders*, *193*, 282–288. https://doi.org/10.1016/j.jad.2016.01.007

Lazarov, A., Ben-Zion, Z., Shamai, D., Pine, D. S., & Bar-Haim, Y. (2018). Free viewing of sad and happy faces in depression: A potential target for attention bias modification. *Journal of Affective Disorders*, *238*, 94–100. https://doi.org/10.1016/j.jad.2018.05.047

Lazarov, A., Pine, D. S., & Bar-Haim, Y. (2017). Gaze-contingent music reward therapy for social anxiety disorder: A randomized controlled trial. *American Journal of Psychiatry*, *174*(7), 649–656. https://doi.org/10.1176/appi.ajp.2016.16080894

Lecrubier, Y., Sheehan, D., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., ... Dunbar, G. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, *12*(5), 224–231. https://doi.org/10.1016/S0924-9338(97)83296-8

Linetzky, M., Kahn, M., Lazarov, A., Pine, D. S., & Bar-Haim, Y. (2019). Gaze-Contingent Music Reward Therapy for Clinically Anxious 7-to 10-Year-Olds: An Open Multiple Baseline Feasibility Study. *Journal of Clinical Child & Adolescent Psychology*, 1-8.

Linetzky, M., Pergamin-Hight, L., Pine, D. S., & Bar-Haim, Y. (2015). Quantitative evaluation of the clinical efficacy of attention bias modification treatment for anxiety disorders. *Depression and Anxiety*, *32*(6), 383–391. https://doi.org/10.1002/da.22344

Maratos, A., Gold, C., Wang, X., & Crawford, M. (2008). Music therapy for depression. *Cochrane Database of Systematic Reviews*(1). https://doi.org/10.1002/14651858.CD004517.pub2

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, *134*(4), 382–389. https://doi.org/10.1192/bjp.134.4.382

Morgan, J. K., Shaw, D. S., Olino, T. M., Musselman, S. C., Kurapati, N. T., & Forbes, E. E. (2016). History of depression and frontostriatal connectivity during reward processing in late adolescent boys. *Journal of Clinical Child & Adolescent Psychology*, *45*(1), 59–68. https://doi.org/10.1080/15374416.2015.1030753

Müller, M. J., Szegedi, A., Wetzel, H., & Benkert, O. (2000). Moderate and severe depression: Gradations for the Montgomery–Åsberg Depression Rating Scale. *Journal of Affective Disorders*, *60*(2), 137–140. https://doi.org/10.1016/S0165-0327(99)00162-7

Nestler, E. J., & Carlezon, W. A. (2006). The mesolimbic dopamine reward circuit in depression. *Biological Psychiatry*, *59*(12), 1151–1159. https://doi.org/10.1016/J.BIOPSYCH.2005.09.018

Newman, K., Quigley, L., Fernandez, A., Dobson, K., & Sears, C. (2019). Concurrent and prospective relations between attentional biases for emotional images and relapse to depression. *Cognitive Therapy and Research*, 1–17. https://doi.org/10.1007/s10608-019-10017-y

Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and Anxiety*, *27*(12), 1135–1142. https://doi.org/10.1002/da.20755

Posternak, M. A., Solomon, D. A., Leon, A. C., Mueller, T. I., Shea, M. T., Endicott, J., & Keller, M. B. (2006). The naturalistic course of unipolar major depression in the absence of somatic therapy. *The Journal of Nervous and Mental Disease*, *194*(5), 324–329. https://doi.org/10.1097/01.nmd.0000217820.33841.53

Rief, W., Nestoriuc, Y., Weiss, S., Welzel, E., Barsky, A. J., & Hofmann, S. G. (2009). Meta-analysis of the placebo response in antidepressant trials. *Journal of Affective Disorders*, *118*(1-3), 1–8. https://doi.org/10.1016/j.jad.2009.01.029

Rotnitzky, A., & Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, *77*(3), 485–497. https://doi.org/10.1093/biomet/77.3.485

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS One*, *4*(10), e7487. https://doi.org/10.1371/journal.pone.0007487

Sheehan, D., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Weiller, E., Keskiner, A., ... Dunbar, G. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, *12*(5), 232–241. https://doi.org/10.1016/S0924-9338(97)83297-X

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., ... Dunbar, G. C. (1998). 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. *The Journal of clinical psychiatry*, *59*(Suppl 20), 22–33.

Soltani, S., Newman, K., Quigley, L., Fernandez, A., Dobson, K., & Sears, C. (2015). Temporal changes in attention to sad and happy faces distinguish currently and remitted depressed individuals from never depressed individuals. *Psychiatry Research*, *230*(2), 454–463. https://doi.org/10.1016/j.psychres.2015.09.036

Sprinkle, S. D., Lurie, D., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., & Bissada, N. N. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *Journal of Counseling Psychology*, *49*(3), 381–385. https://doi.org/10.1037/0022-0167.49.3.381

Vanderhasselt, M. A., & De Raedt, R. (2009). Impairments in cognitive control persist during remission from depression and are related to the number of past episodes: An event related potentials study. *Biological Psychology*, *81*(3), 169–176. https://doi.org/10.1016/j.biopsycho.2009.03.009

Webb, C. A., Kertz, S. J., Bigda-Peyton, J. S., & Björgvinsson, T. (2013). The role of pretreatment outcome expectancies and cognitive-behavioral skills in symptom improvement in an acute psychiatric setting. *Journal*

of *Affective Disorders*, *149*(1–3), 375–382. https://doi.org/10.1016/j.jad.2013.02.016

Whiteford, H. A., Harris, M. G., Mckeon, G., Baxter, A., Pennell, C., Barendregt, J. J., & Wang, J. (2013). Estimating remission from untreated major depression: A systematic review and meta-analysis. *Psychological Medicine*, *43*, 1569–1585. https://doi.org/10.1017/S0033291712001717

Whitton, A. E., Treadway, M. T., & Pizzagalli, D. A. (2015). Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Current Opinion in Psychiatry*, *28*(1), 7–12. https://doi.org/10.1097/YCO.0000000000000122

Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*(1), 121–130. https://doi.org/10.2307/2531248

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, *44*(4), 1049–1060.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.