



Published in final edited form as:

Clin Psychol Sci. 2020 July ; 8(4): 756–772. doi:10.1177/2167702620902119.

Randomized Controlled Trial Testing Mobile-Based Attention-Bias Modification for Posttraumatic Stress Using Personalized Word Stimuli

Andrea N. Niles^{1,2}, Joshua D. Woolley^{1,2}, Paige Tripp^{1,2}, Ana Pesquita^{1,3}, Sophia Vinogradov⁴, Thomas C. Neylan^{1,2}, Aoife O'Donovan^{1,2}

¹San Francisco Veterans Affairs Medical Center, San Francisco, California;

²Department of Psychiatry, University of California, San Francisco;

³Center for Human Brain Health, University of Birmingham;

⁴Department of Psychiatry, University of Minnesota

Abstract

Although behavioral therapies are effective for posttraumatic stress disorder (PTSD), access for patients is limited. Attention-bias modification (ABM), a cognitive-training intervention designed to reduce attention bias for threat, can be broadly disseminated using technology. We remotely tested an ABM mobile app for PTSD. Participants ($N = 689$) were randomly assigned to personalized ABM, nonpersonalized ABM, or placebo training. ABM was a modified dot-probe paradigm delivered daily for 12 sessions. Personalized ABM included words selected using a recommender algorithm. Placebo included only neutral words. Primary outcomes (PTSD and anxiety) and secondary outcomes (depression and PTSD clusters) were collected at baseline, after training, and at 5-week-follow-up. Mechanisms assessed during treatment were attention bias and self-reported threat sensitivity. No group differences emerged on outcomes or attention bias. Nonpersonalized ABM showed greater declines in self-reported threat sensitivity than placebo ($p = .044$). This study constitutes the largest mobile-based trial of ABM to date. Findings do not support the effectiveness of mobile ABM for PTSD.

Keywords

attention; evidence-based treatments; posttraumatic stress disorder

Article reuse guidelines: sagepub.com/journals-permissions

Corresponding Author: Aoife O'Donovan, University of California, San Francisco, and San Francisco VA Medical Center, 4150 Clement St., 116-C1, San Francisco, CA 94121, aoife.odonovan@ucsf.edu.

Author Contributions

A. O'Donovan developed the study concept. A. O'Donovan, J. D. Woolley, and A. N. Niles contributed to the study design. Testing and data collection were performed by A. N. Niles and P. Tripp, and A. N. Niles performed the data analysis and interpretation under the supervision of A. O'Donovan and J. D. Woolley. A. N. Niles drafted the manuscript, and all of the other authors provided critical revisions. All of the authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2167702620902119>

Posttraumatic stress disorder (PTSD) is a highly distressing and disabling condition accounting for significant health burden worldwide (Kessler et al., 2017). Although effective treatments are available for PTSD, such as prolonged exposure (Powers, Halpern, Ferenschak, Gillihan, & Foa, 2010) and cognitive processing therapy (Cusack et al., 2016), these treatments are costly, time-consuming, not available to all patients, and dependent on patients frequently attending the clinic for services. Seventy-seven percent of people in the United States own smartphones (Pew Research Center, 2019), which means that treatments administered on mobile devices have the potential to solve challenges to treatment access. PTSD has been linked with elevated sensitivity to threat as indexed by altered affective and biological reactivity to threat-related stimuli (Badura-Brack et al., 2018; Niles et al., 2018; Vasterling & Arditte Hall, 2018) and with threat-related attention biases, in some but not all studies (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Fani et al., 2012; Kruijt, Putman, & Van der Does, 2013). In fact, exaggerated threat sensitivity may underlie several symptoms of PTSD, including alterations in arousal, reactivity, and avoidance. Attention-bias modification (ABM), which uses implicit cognitive training strategies to orient attention allocation away from threatening stimuli, can easily translate to a mobile device and holds promise for reducing symptom severity and enhancing treatment access among individuals with PTSD. In the present study, we examined the effectiveness of an ABM mobile app for people with elevated PTSD symptoms.

The procedure for ABM involves presentation of two stimuli on a screen—one neutral and one threatening. Stimuli typically used are either faces or words. The stimuli are presented for a brief period (typically 500 ms) before they disappear, and a probe appears in place of the neutral stimulus. The participant then identifies some aspect of the probe (e.g., direction, type). This process is repeated over many trials. The goal of this training is to encourage attention allocation away from threat stimuli and toward neutral stimuli during early information processing. A review of meta-analyses indicated that ABM is effective for reducing anxiety symptoms in both trait and clinically anxious samples (Jones & Sharpe, 2017).

Compared with the large number of ABM studies conducted for individuals with social anxiety disorder, relatively few studies have examined effects of ABM in PTSD. Given that hypervigilance for threat is one of the defining characteristics of PTSD and that ABM is hypothesized to act on attentional processes that maintain threat hypervigilance, ABM could be particularly beneficial for this population. Results are mixed, however, for clinical trials examining ABM in PTSD. One trial showed that ABM outperformed control training (Kuckertz, Amir, et al., 2014), another showed that control training outperformed ABM (Badura-Brack et al., 2015), and two additional trials found no difference between ABM and control training for symptom reduction (Khanna et al., 2016; Schoorl, Putman, & Van Der Does, 2013). In a recent review, Woud, Verwoed, and Krans (2017) discussed the next steps for ABM in PTSD and suggested that more randomized clinical trials are needed given the paucity of research in this area. Thus, we aimed to examine the effectiveness of ABM in participants with elevated PTSD symptoms in a large-scale, remote, randomized clinical trial. Furthermore, because attention bias away from threat is the theorized mechanism targeted by ABM, one possible explanation for prior null findings is that ABM is more effective for those showing a baseline attention bias and that the noise introduced by

baseline variability in attention bias had obscured significant effects. Indeed, some prior studies have shown that greater baseline attention bias predicts greater symptom reduction following ABM (Amir, Taylor, & Donohue, 2011; Kuckertz, Gildebrant, et al., 2014). However, some studies have indicated that ABM works in the absence of effects of attention bias (Badura-Brack et al., 2015), and a recent meta-analysis indicated a general lack of attention bias at baseline in clinically anxious individuals (including those with PTSD) enrolled in ABM trials (Kruijt, Parsons, & Fox, 2019). Nonetheless, we further examined whether baseline attention bias predicted response to ABM.

In the control condition used in prior studies, called *attention-control training*, the probe appears in place of the neutral stimulus 50% of the time and in place of the threat stimulus 50% of the time. Thus, participants are not trained to orient attention away from threat, but the control condition is matched to the active treatment for the presentation of threatening stimuli and the type of activity completed. Badura-Brack and colleagues (2015) theorized, on the basis of their unexpected finding that the control outperformed ABM, that attention-control training may be effective for the treatment of PTSD by encouraging greater attentional flexibility in the context of threat regardless of threat-related contingencies for probe location. Thus, in an effort to maximize group differences, we compared ABM with a control training condition that included only neutral words to test whether training users to allocate attention away from threat resulted in better outcome compared with a non-threat-related attention task.

The ultimate impact of computerized cognitive training programs would be much greater if they can be administered remotely, at low cost, and without the assistance of a clinician. However, prior studies examining remote administration of ABM either via mobile phone or computer have generally demonstrated no benefit of ABM over attention-control training (Linetzky, Pergamin-Hight, Pine, & Bar-Haim, 2015). Prior studies have been conducted among participants with social anxiety symptoms (Boettcher, Berger, & Renneberg, 2012; Carlbring et al., 2012; Neubauer et al., 2013) and nonclinical samples (with encouragement of participation by those with social or generalized anxiety symptoms; Enock, Hofmann, & McNally, 2014). All remote studies have demonstrated significant within-group symptom reduction in both the ABM and attention control conditions. Thus, it is unclear from these findings whether ABM is ineffective when administered remotely or both ABM and attention-control training may effectively target anxiety-related attentional processes to reduce anxiety symptoms. Furthermore, it is not clear whether remote ABM could be effective for people with elevated PTSD symptoms.

The majority of ABM studies have used face stimuli in ABM training. Faces are relevant for a sample with social anxiety symptoms because this population is likely to find these stimuli particularly threatening. However, for other anxiety presentations, facial stimuli may be less relevant. Thus, previous studies have also used words in ABM. A recent review of meta-analyses of cognitive bias modification studies concluded that ABM employing word stimuli is more effective at reducing anxiety symptoms compared with studies using face stimuli (Jones & Sharpe, 2017). Researchers have attempted to augment the effects of ABM by using word stimuli that are personally relevant to each individual participant. Although no prior study has directly compared personalized ABM with nonpersonalized

ABM, various methods have been used to personalize the stimuli used in training, including having participants rate stimuli (Amir, Beard, Burns, & Bomyea, 2009; Amir, Kuckertz, & Strege, 2016) or transcription of diagnostic interviews (Schoorl, Putman, Mooren, Van Der Werff, & Van Der Does, 2014). Because the goal of ABM is to improve a person's ability to disengage attention from threat, using stimuli that are particularly threatening should increase the task demand and theoretically augment the effectiveness of the training. Niles and O'Donovan (2018) described a novel method for automating the selection of personalized word stimuli in PTSD that uses a recommender-algorithm machine-learning approach. One goal of the present study was to test whether using personalized word stimuli, selected using this recommender algorithm, improves ABM effectiveness.

We compared the effects of personalized ABM and nonpersonalized ABM with that of placebo attention training (including only neutral stimuli) administered entirely remotely on participants' mobile devices using a mobile app called "Resolving Psychological Stress" (RePS). Our first aim was to examine treatment-group differences in symptom change. We hypothesized that participants assigned to personalized ABM would show greater symptom reduction than those assigned to nonpersonalized ABM and placebo training and that participants assigned to nonpersonalized ABM would show greater symptom reduction than those assigned to placebo training. Our second aim was to test group differences in the change in putative mechanisms for ABM, including attention bias for threat and self-reported threat sensitivity. We hypothesized that participants assigned to personalized ABM would show greater change in attention bias for threat and threat sensitivity than participants assigned to nonpersonalized ABM and placebo training and that participants assigned to nonpersonalized ABM would show greater change in these putative ABM mechanisms than participants assigned to placebo training. Our third aim was to examine whether baseline attention bias was a moderator of treatment response. We hypothesized that participants showing more attention bias at baseline would benefit more from ABM than from placebo training.

Method

Design

Participants were randomly assigned to one of three training conditions (personalized ABM, nonpersonalized ABM, and placebo), and assignment was stratified by gender and symptom severity on the PTSD Checklist for *DSM-5* (PCL-5) assessed during the eligibility screen (low = 33–49, medium = 50–65, high = 66–80; Blevins, Weathers, Davis, Witte, & Domino, 2015). Randomization was determined using a random-number-generator function in Microsoft Excel 2016, and the research coordinator created the allocation scheme and assigned participants to conditions. The research coordinator was not blind to condition assignment, but all other study personnel were blind until data collection was complete. Symptoms were assessed before beginning ABM training (baseline), directly after completion of training (posttraining), and at 5 weeks after the start of training (follow-up). Assessment of ABM mechanisms occurred during training. Attention bias was assessed at baseline, halfway through completion of training (midtraining) and at posttraining, and self-reported threat sensitivity was assessed on training Sessions 2 through 7 and 9 through

14 (for a total of 12 assessment points). The study was approved by the Institutional Review Board at the University of California, San Francisco. Hypotheses and the planned analytic approach were preregistered internally within the lab and approved by all authors after commencing data collection but before conducting any analyses or unblinding of researchers to condition assignment.¹

Participants

Participants were recruited via Craigslist ads posted in cities across the United States and ads posted to Reddit discussion forums dedicated to posttraumatic stress or specific trauma groups (e.g., survivors of sexual assault). Advertisements were posted in every state, and participants were enrolled from every state, including Washington, D.C., with the exception of Mississippi and Wyoming. Advertisements directed participants to an online screening questionnaire (administered via Qualtrics) with which they were assessed for eligibility. Eligible participants were over 18 years old, owned an iPhone or iPod Touch running iOS 10, and scored 33 or above on the PCL-5. A cutoff score of 33 was based on recommendations by the National Center for PTSD (2018) to indicate probable PTSD diagnosis. Participants reporting active suicidal ideation were excluded and directed to a website with resources related to suicide. Those completing the online screening multiple times were also excluded. As shown in the Consolidated Standards of Reporting Trials (CONSORT) diagram in Figure 1, 2,627 participants completed the eligibility screening, 1,694 participants were eligible to participate in the study, and 1,001 participants confirmed interest and were officially enrolled in the study and randomized to conditions. Of those participants, 689 completed the baseline assessment on the mobile app and thus were included in analyses. Demographic and clinical characteristics are reported in Table 1.

Procedure

Details of the full study protocol can be requested from A. O'Donovan via e-mail. Eligible participants were asked to provide electronic consent to participate in the study and were sent a follow-up e-mail to confirm interest in participating. Those who confirmed interest were randomly assigned to one of the three training conditions and were sent a link to complete the baseline questionnaires (also administered via Qualtrics). After completing the baseline questionnaires, participants were sent instructions on how to access and download RePS.

Participants were given 21 days to complete 15 activity sessions on RePS. Each session, participants responded to questions on the app and completed either the dot-probe assessment or ABM training. Participant engagement with RePS was monitored via a back-end portal. Participants who had not engaged with the app for 3 days were sent an e-mail requesting that they reengage. If participants still had not engaged after 6 days, they were sent a second e-mail requesting that they reengage or notify the experimenter if they wanted to withdraw from the study. If participants did not complete the 15 days of training within the 21-day window, RePS automatically pushed to the final day, which

¹This preregistration can be obtained from A. O'Donovan. Because of a miscommunication between A. N. Niles and A. O'Donovan when the latter was on leave, the preregistration document was never formally uploaded to any online system.

included questionnaires and the final dot-probe assessment, and the experimenter prompted the participant to complete this final assessment. We encouraged participants to complete this final assessment regardless of whether they completed all ABM training sessions.

After completion of the final assessment on the app, participants were sent a link to complete the posttraining questionnaires online via Qualtrics. Five weeks after the first day of engagement with RePS, participants were sent a link to complete a final set of questionnaires online. Participants received \$2 for the baseline and posttraining questionnaire assessments, \$5 for the follow-up assessment, \$1.50 per day for Sessions 1 through 14 of the app, and \$2 for Session 15 of the app for a total possible compensation of \$35. Recruitment and data collection, which began in January 2018 and were completed in October 2018, were managed by one full-time research coordinator and a postdoctoral fellow.

Treatments

ABM was administered via the RePS app on participants' mobile devices. Training occurred in Sessions 2 through 7 and Sessions 9 through 14 for a total of 12 sessions of ABM training. For ABM, following procedures used in previous studies (Badura-Brack et al., 2015; Schoorl et al., 2014), participants first viewed a fixation cross for 500 ms, then two words appeared located 3 cm apart for 500 ms. Words appeared in 4-mm high capital letters in the SF UI Display Light typeface. Assuming that the average user holds the phone at a distance of 20 to 50 cm, then the stimuli occupied 0.5 to 1.5 degrees of visual angle, which is comparable with the stimuli size in previous computerized versions of the ABM task (Badura-Brack et al., 2015; Najmi & Amir, 2010). The words then disappeared, and a probe letter (either an "E" or an "F") appeared in place of one of the words. Participants then selected which letter appeared by pressing the corresponding button at the bottom of the screen. The letters stayed on the screen until the participant provided a response (for examples of how training appeared on the app, see Fig. S1 in the Supplemental Material available online). Font size and word distance were fixed such that they would appear the same way regardless of phones' screen size. Participants assigned to the nonpersonalized and personalized ABM conditions viewed one threat word and one neutral word, and the probe appeared in place of the neutral word 100% of the time. Threat-word stimuli used in the personalized group were selected to be personally relevant for each individual (i.e., each person in this group had a different personalized word set), whereas in the nonpersonalized group, all participants had the same set of threat-word stimuli (i.e., stimuli were not personalized to the individual user). Participants in the placebo condition viewed two neutral words instead of one threat word and one neutral word. Each training session included 90 trials, which is comparable with previous ABM trials (Badura-Brack et al., 2015; Schoorl et al., 2014). The app was programmed to restart the ABM training for the day if the participant left the app or answered a call or text message during the training.

Word stimuli

Threat-relevant words were culled from multiple studies in which word stimuli were selected for PTSD populations (Beck, Freeman, Shipherd, Hamblen, & Lackner, 2001; Blix & Brennen, 2011; Bunnell, 2007; Ehring & Ehlers, 2011; Roberts, Hart, & Eastwood, 2010;

Swick, 2009) as well as from the Affective Norms for English Words (ANEW) standard word rating set (Bradley & Lang, 1999) and a study examining cognitive processing of emotional words (Kousta, Vinson, & Vigliocco, 2009). Duplicate words and words that may have been familiar only to military personnel (e.g., IED, Falluja, APC, Kirkuk) were removed, which resulted in a final set of 513 threat words. Words were then rated online using Amazon Mechanical Turk (for details, see Niles & O'Donovan, 2018); 1,112 participants who reported having experienced a traumatic event rated each word on a Likert scale from 1 (*most threatening*) to 9 (*least threatening*). Sixty-four words with average threat ratings greater than 4 were reserved for the dot-probe assessment used to measure attentional bias for all participants across the three conditions. Thus, 453 words remained for use in ABM training.

Personalized word selection.—For the personalized ABM training, a recommender system was used to select a personalized set of 60 words from the pool of 453 remaining words after removal of those used in the assessment (for details, see Niles & O'Donovan, 2018). All participants rated 55 words when first opening the RePS app. For participants assigned to the personalized condition, ratings were then used to select the personalized word set used in training. The recommender system used the initial 55 word ratings to find users in the training data set (described above) who had similar answers. The recommender system used these similarities to make predictions about the threat level that the target user would assign to all 453 words. The 60 words with the highest predicted ratings were then selected for use in that individual's ABM training. As discussed in Niles and O'Donovan (2018), the recommender algorithm identified a set of words with a threat rating that was 91% of the maximum possible threat rating for each user (compared with 61% for a random set of threat words). The difference in accuracy between the algorithm-selected and randomly selected words was highly significant with an effect size (Cohen's *d*) of 2.92.

The 60 selected words were then subdivided into two lists of 30 words; the top 30 most threatening words were in one list, and the next 30 most threatening words were in a second list. Within the ABM training, the first week of training (Sessions 2–7) used the less threatening word set, and the second week of training (Sessions 9–14) used the more threatening word set. This procedure is akin to an exposure hierarchy in cognitive behavioral therapy and was intended to “step up” the challenge of ABM training from the first to the second week.

Nonpersonalized-word selection.—Sixty words were randomly selected for use in nonpersonalized ABM training. All participants in this condition had the same set of words. Words were randomly divided into two lists of 30 words; one list was administered during the first week of training, and the second list was administered during the second week of training.

Neutral words and placebo word selection.—Neutral words were selected from neutral words used in previous studies and the ANEW word list. All 513 threat words were matched with a neutral word of the same length, and additional neutral words were needed for the placebo condition. Words from previous literature did not provide enough matches for all words in our word pool. Thus, we selected additional neutral words using

the dictionary and searches for specific word lengths in Google. We primarily selected nouns and verbs that were not associated with trauma or violence or with positive emotions or achievement. Words were selected by a research assistant and were then reviewed by two doctoral-level researchers and two additional research assistants to confirm neutrality. We then used the *Corpus of Contemporary American English* (Davies, 2008) to identify the frequency with which each neutral and threat word was used in the English language. Threatening and neutral words were then matched in terms of frequency of use and word length. Participants assigned to the placebo condition viewed only pairs of neutral words during ABM training.

Measures

Trauma exposure.—Trauma exposure was assessed using the Trauma History Screen (THS; Carlson et al., 2011). Participants were asked whether they had experienced 14 different traumatic events and were asked to identify the number of times each endorsed event had occurred. Participants were then asked to categorize the worst event into 1 of the 14 THS categories and identify the age when the event occurred. Trauma exposure of the sample is described in Table S1 in the Supplemental Material (similar categories on the THS were combined for parsimony).

Treatment outcome measures.—Primary outcomes were the PCL-5 and the State Trait Anxiety Inventory Trait Version (STAI). Secondary outcomes were the four subscales of the PCL-5 (intrusion, avoidance, negative cognitions/mood, and arousal/reactivity) and the depression subscale of the Depression Anxiety and Stress Scale (DASS).

PCL-5.: The 20-item PCL assesses for the 20 symptoms of PTSD included in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5;* American Psychiatric Association, 2013) and is a widely used and well-validated measure of PTSD symptom severity (Blevins et al., 2015). Participants rated how much they have been bothered by each symptom on a scale from 0 (*not at all*) to 4 (*extremely*). For the present study, participants were asked to consider their symptoms within the past week. We calculated a total symptom score across all 20 scale items as well as cluster severity scores for clusters B (intrusion), C (avoidance), D (negative cognitions/mood), and E (arousal/reactivity) of the *DSM-5* criteria. In the present study, for the full scale, α values were .92, .96, and .95 for baseline, posttraining, and follow-up, respectively.

STAI-Trait Version.: The 20-item STAI (Spielberger, 1985) assesses for anxiety symptom severity. Participants rate the frequency of experiencing each item on a 1 (*almost never*) to 4 (*almost always*) Likert scale. In the present study, α values were .90, .90, and .91 for baseline, posttraining, and follow-up, respectively. The STAI was included as a primary outcome measure because it has been used frequently in prior studies of ABM (Cristea, Kok, & Cuijpers, 2015).

DASS Depression subscale.: The seven-item Depression subscale of the DASS (Antony, Bieling, Cox, Enns, & Swinson, 1998) assesses symptoms of dysphoric mood. For the present study, participants were asked to consider their symptoms within the past week. The

measure has acceptable to excellent internal consistency and concurrent validity (Antony et al., 1998). In our study, α values were .89, .90, and .92 for baseline, posttraining, and follow-up, respectively.

Treatment-mechanism measures.—To assess theorized mechanisms of the effects of ABM on anxiety reduction, we included dot-probe assessments to measure attention bias for threat and measures of self-reported threat sensitivity during treatment.

Attention bias.: Attention bias was assessed using the dot-probe paradigm administered on the RePS app on participants' mobile devices. Dot-probe assessments were administered in Sessions 1, 8, and 15 of the RePS program. During dot-probe assessments, participants first viewed a fixation cross for 500 ms. The cross then disappeared, and two words, one threat and one neutral word, appeared on the top and bottom of the screen. The words were 3 cm apart and remained on the screen for 500 ms. The words then disappeared, and a letter, either an "E" or an "F," appeared in place of one of the words. The participant then used his or her thumbs to press a response at the bottom of the screen to indicate whether the letter was an "E" or an "F." For an example of how the assessment appeared on the app, see Figure S1 in the Supplemental Material. The probe remained on the screen until the participant provided a response. The probe appeared in place of the neutral word (incongruent trials) 50% of the time and in place of the threat word (congruent trials) 50% of the time. Threat word location, probe location, and probe type ("E" or "F") were counterbalanced and randomized across all trials. Participants completed a total of 70 trials, and the first 6 trials were treated as practice trials. Font type and size were the same as in ABM training.

Threat-sensitivity scale.: Threat sensitivity was assessed using an eight-item measure (developed for use in the present study) consisting of items from multiple different scales, including the Attentional Control Scale (three items; Derryberry & Reed, 2002), the Threat Orientation Scale (three items; Thompson, Schlehofer, & Bovin, 2006), and the Arousal/Reactivity subscale of the PCL-5 (two items; Blevins et al., 2015). All scale items are shown in Table S2 in the Supplemental Material. Participants were asked about the previous 24 hr for each item and responded on a Likert scale from 1 (*not at all*) to 5 (*extremely*). These items were selected by the researchers to capture the hypothesized mechanisms of ABM, including attentional control, perception of acute threat, and hypervigilance. Across all sessions for all participants, α was .93.

Statistical analyses

Attention-bias data cleaning.—Because of the remote administration of the dot-probe task and thus the uncontrolled nature of its administration, attention-bias data were carefully inspected before making decisions regarding outlier removal. Before calculating the mean reaction time on the task, we removed incorrect trials, trials shorter than 300 ms, and trials longer than 1,500 ms. These thresholds were based on visual observation of the distributions of our data and those in prior research (Price et al., 2015). The average percentage of incorrect trials was 2% ($SD = 3$; range = 0%–20%) in Sessions 1, 8, and 15. We also examined the distribution of counts for incorrect trials, trials longer than 1,500 ms, and trials shorter than 300 ms for each dot-probe assessment session. We removed reaction-time data

in sessions that were above the 99th percentile for these counts (more than 18 incorrect trials in a session, more than 26 trials shorter than 1,500 ms in a session, and more than 4 trials longer than 300 ms in a session) because these extreme values likely indicate that the participant was not engaging with the assessment as instructed for that session. Percentages of participants removed from analyses were 4% ($n = 31$) in Session 1, 3% ($n = 19$) in Session 8, and 5% ($n = 24$) in Session 15.

Treatment effects on symptom reduction.—Primary symptom outcomes included the PCL-5 and the STAI, and secondary outcomes were the PCL-5 subscales and the Depression subscale of the DASS. Data were nested within individuals as a result of repeated measurement at multiple time points, so we used multilevel modeling (MLM), which allow for estimation of within-subject and between-subject effects. MLM also accommodates unequal numbers of observations across individuals, which allows the inclusion of participants who are missing one or more assessments in their entirety. Thus, we conducted an intent-to-treat analysis, including all participants who completed the baseline assessment on the RePS app ($N = 689$). Models included the Time \times Condition interaction as well as main effects of time and condition, and we examined the Time \times Condition interaction effect for statistical significance. Because of a pattern of symptom change typically observed in clinical trials in which the largest change occurred from pretraining to posttraining, with a leveling out of change from posttraining through follow-up, effects from pretraining to posttraining and from posttraining to follow-up were examined using a piecewise approach, which has been used previously to examine data from clinical trials (Craske et al., 2014; Niles, Craske, Lieberman, & Hur, 2015; Roy-Byrne et al., 2010; Wolitzky-Taylor et al., 2018). Thus, the posttraining time point was the reference group in all models. For condition, because of a priori hypotheses, we examined all pairwise condition comparisons (personalized vs. nonpersonalized, personalized vs. placebo, and nonpersonalized vs. placebo) regardless of whether the omnibus Time \times Condition interaction was statistically significant. For primary and secondary outcomes, we also tested within-group effects of time. Inclusion of random effects for slopes and covariance between slope and intercept random effects did not improve model fit when models were compared using likelihood ratio tests. Thus, final models included only random intercepts. Effect sizes are calculated for all Time \times Condition interactions according to the method described by Feingold (2009), which produces estimates analogous to Cohen's d for growth curve models in randomized clinical trials.

Treatment effects on purported mechanisms.—We then tested whether change in purported mechanisms differed by treatment group. We examined change in attention bias and self-reported threat sensitivity as mechanisms. For the analysis of attention bias, we examined congruence (congruent vs. incongruent) as an additional within-subjects factor. Thus, models examining attention bias included reaction time as the dependent measure predicted from the three-way Time \times Condition \times Congruence interaction (as well as all lower order two-way interactions and main effects). The three-way interaction was examined for statistical significance. Times included baseline, midtraining, and posttraining (we did not assess attention bias at follow-up) and was treated as a continuous (linear) effect in the model. Again, regardless of statistical significance of the omnibus test, we examined

all three pairwise comparisons between the groups for statistical significance. Models included random intercepts, slopes (time), and the covariance between intercept and slope because inclusion of these random effects significantly improved model fit when tested using likelihood ratio tests.

For the analysis of self-reported threat sensitivity, models included threat sensitivity as the dependent measure predicted from the two-way Time \times Condition interaction and main effects of time and condition. The two-way interaction was examined for statistical significance. Time included training Sessions 1 through 12 and was treated as a continuous (linear) effect in the model. All three pairwise comparisons between groups were tested for statistical significance. Models included random intercepts, slopes (time), and the covariance between intercept and slope because inclusion of these random effects significantly improved model fit when tested using likelihood ratio tests. All analyses were conducted in Stata 15.

Power analysis

We used G*Power (Version 3.1.9; Faul, Erdfelder, Lang, & Buchner, 2007) to conduct a power analysis before data collection. To achieve 80% power to detect pairwise group differences with a small effect (Cohen's $d = 0.30$) using an independent-samples t test, 176 participants were needed per group for a total sample size of 528 participants. Posttraining data were collected from 546 participants (personalized $n = 177$, nonpersonalized $n = 179$, placebo $n = 190$), and the trial was stopped because the necessary sample size was achieved. Because data were analyzed using a random effects model rather than a t test, achieved power to detect a small effect for symptom reduction was 99%, and we had 80% power to detect a very small effect size (Cohen's $d = 0.15$).

Results

Pretraining group differences

In the intent-to-treat sample ($N = 689$), we examined baseline group differences on primary and secondary outcomes and purported mechanisms. No group differences emerged for PCL-5 ($p = .953$), STAI ($p = .965$), PCL-5 subscales (p s $> .302$), DASS Depression ($p = .865$), attention bias ($p = .590$), or threat sensitivity ($p = .664$).

Treatment adherence and attrition

Of the 1,001 participants who confirmed interest in participating in the study, 689 (68.8%) downloaded and completed the first session (i.e., baseline attention bias and PCL-5 assessment) of RePS. Only participants who completed this baseline assessment were included in analyses because at least one PCL assessment was needed for inclusion. No significant differences were found in initial engagement among the three treatment groups, $\chi^2(2, N = 1001) = 0.27, p = .875$. Of the 689 participants who completed the baseline assessment on the app, 20 (2.9%) did 0 training sessions, 116 (17%) completed between 1 and 6 training sessions, 70 (10%) completed between 7 and 11 sessions of training, and 483 (70.1%) completed all 12 sessions (see consort for number of completers by session). No

significant difference was found in attrition (defined as completion of fewer than 10 training sessions) between the three treatment groups, $\chi^2(2, N = 689) = 1.89, p = .389$.

Treatment effects on symptom reduction

Primary outcomes.—Primary outcomes included the PCL-5 and the STAI. Parameter estimates and effect sizes for the Time \times Condition interaction are shown in Table 2, and estimated means by condition at the three time points are shown in Table S3 in the Supplemental Material.

For the PCL, the omnibus test for the Time \times Condition interaction was not significant ($p = .786$). Regarding a priori contrasts, for the comparison of personalized ABM with nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Time \times Condition interactions emerged from baseline to posttraining ($ps > .524$) or from posttraining to follow-up ($ps > .525$). All three groups showed a significant reduction in PCL scores from baseline to posttraining ($ps < .001$) but not from posttraining to follow-up ($ps > .484$).

For the STAI, the omnibus test for the Time \times Condition interaction was not significant ($p = .398$). For the comparison of personalized ABM with nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Time \times Condition interactions emerged from baseline to posttraining ($ps > .089$) or from posttraining to follow-up ($ps > .339$). The personalized group showed a significant reduction in STAI scores from baseline to posttraining ($p = .003$), whereas the nonpersonalized and placebo groups did not ($ps > .280$). From posttraining to follow-up, the placebo group showed a significant reduction in STAI scores ($p = .021$), whereas the other two groups did not ($ps > .355$).

Secondary outcomes.—Secondary outcomes included the subscales of the PCL and the Depression subscale of the DASS. Parameter estimates and effect sizes for the Time \times Condition interaction are shown in Table 2, and estimated means by condition at the three time points are shown in Table S1 in the Supplemental Material. For the subscales of the PCL, none of the omnibus tests for the Time \times Condition interactions were significant ($ps > .269$). Regarding a priori contrasts, for the comparison of personalized ABM with nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Time \times Condition interactions emerged from baseline to posttraining ($ps > .200$) or from posttraining to follow-up ($ps > .291$). All three groups showed a significant reduction in all PCL subscale scores from baseline to posttraining ($ps < .001$). For avoidance, the placebo group showed a significant increase from posttraining to follow-up ($p = .001$), whereas the other groups did not ($ps > .06$). For negative cognitions and mood, the nonpersonalized ABM group showed a significant reduction from posttraining to follow-up ($p = .046$), whereas the other groups did not ($ps > .134$). No significant changes from posttraining to follow-up were found for the Intrusion or Arousal/Reactivity subscales of the PCL for any group ($ps > .258$).

For the DASS Depression subscale, the omnibus test for the Time \times Condition interaction was not significant ($p = .543$). Regarding a priori contrasts, for the comparison of

personalized ABM with nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Time \times Condition interactions emerged from baseline to posttraining ($p > .256$) or from posttraining to follow-up ($p > .397$). None of the groups showed a significant reduction in DASS depression scores from baseline to posttraining ($p > .069$). The nonpersonalized and placebo groups showed a significant reduction in DASS depression scores from posttraining to follow-up ($p < .012$), whereas the personalized group did not ($p = .548$).

Sensitivity analyses.—For primary outcomes, we conducted two follow-up analyses on subsets of participants to determine whether group differences emerged for people who received a greater dose of ABM and participants demonstrating greater engagement with RePS. Because participants completed the training remotely, we had little control over environmental factors that could affect attention during training. Thus, to determine whether null findings for group differences could be explained by inclusion of participants who may not have received a sufficient dose of ABM because of insufficient training or inattention during training, we tested group differences for (a) participants who completed at least 10 sessions of ABM and (b) participants who completed at least 10 sessions of ABM and with faster reaction times (bottom 50th percentile) throughout ABM training.

Five hundred thirteen participants completed 10 or more training sessions. For the PCL and STAI, neither of the omnibus tests for the Time \times Condition interaction was significant in this subset ($p = .330$). Regarding a priori contrasts, for the comparison of personalized ABM with nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Time \times Condition interactions emerged from baseline to posttraining ($p > .062$) or from posttraining to follow-up ($p > .221$) for either the PCL or the STAI.

To examine participants with faster reaction times, we removed incorrect trials, trials shorter than 300 ms, and trials longer than 1,500 ms on the training sessions. We also removed data from training sessions on which the user was above the 99th percentile for incorrect trials in a session (more than 13 trials), number of trials shorter than 1,500 ms in a session (more than 24 trials), and number of trials longer than 300 ms in a session (more than 3 trials). We then calculated the average reaction time for each participant across all 15 sessions of RePS and split the data at the median reaction time; we analyzed only those below the 50th percentile and who had completed at least 10 training sessions ($N = 252$). For the PCL and STAI, neither of the omnibus tests for the Time \times Condition interaction was significant ($p > .842$). Regarding a priori contrasts, for the comparison of personalized ABM with nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Time \times Condition interactions emerged from baseline to posttraining ($p > .636$) or from posttraining to follow-up ($p > .269$) for either the PCL or the STAI.

Treatment effects on purported mechanisms

Results for the change in attention bias are shown in Figure S2 in the Supplemental Material. The omnibus test for the Time \times Condition \times Congruent interaction was not significant ($p = .406$). Regarding a priori contrasts, for the comparison of personalized ABM with

nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Time \times Condition \times Congruent interactions emerged ($ps > .190$).

For self-reported threat sensitivity, the omnibus test for the Time \times Condition interaction was not significant ($p = .112$). As shown in Figure 2, for a priori contrasts, the comparison of nonpersonalized ABM with placebo was significant ($b = -0.12$, 95% confidence interval [CI] = $[-0.24, -0.00]$, $p = .044$) such that the slope for threat sensitivity over the 12 training sessions was more negative in nonpersonalized ABM compared with placebo. No additional significant effects emerged for the comparison of personalized ABM with nonpersonalized ABM or the comparison of personalized ABM with placebo ($ps > .142$). Results were unchanged when the threat-sensitivity-scale score was calculated excluding the items from the PCL.

Moderation by baseline attention bias

Finally, we examined whether baseline attention bias (bias; calculated as the difference between incongruent and congruent trials) moderated the effect of treatment condition on PCL-5 and STAI over time. Preliminary analyses showed that baseline attention bias did not differ significantly from 0, $t(652) = 0.6$, $p = .554$, and that baseline attention bias was not associated with PCL-5 total or any PCL-5 subscale ($rs < .03$, $ps > .519$).

For PCL-5, the omnibus test for the Bias \times Condition \times Time interaction was not significant ($p = .535$). Regarding a priori contrasts, for the comparison of personalized ABM with nonpersonalized ABM, nonpersonalized ABM with placebo, and personalized ABM with placebo, no significant Bias \times Time \times Condition interactions emerged ($ps > .140$). For the STAI, the omnibus test for the Bias \times Condition \times Time interaction was not significant ($p = .169$). Regarding a priori contrasts, for the comparison of nonpersonalized ABM with placebo from posttraining to follow-up, the Bias \times Condition \times Time interaction was significant ($b = -0.05$, 95% CI = $[-0.09, -0.00]$, $p = .049$). For the comparison of personalized ABM with placebo from posttraining to follow-up, the Bias \times Condition \times Time interaction was also significant ($b = -0.04$, 95% CI = $[-0.09, -0.00]$, $p = .038$). Tests of simple interaction effects revealed that for users with bias scores 1 *SD* below the mean, STAI scores decreased more so in the placebo group than in the personalized ($b = 2.45$, 95% CI = $[0.17, 4.73]$, $p = .035$) and nonpersonalized ($b = 2.90$, 95% CI = $[0.46, 5.34]$, $p = .020$) groups. No additional significant simple interactions emerged ($ps > .183$).

Dot probe split-half reliability

To calculate the split-half reliability of the dot-probe task, we first examined the correlation between reaction times in the first and second halves of the task. Then, we examined the correlation between the bias score (calculated as the difference between incongruent and congruent trials) in the first and second halves of the task. Reaction times were highly correlated for congruent trials ($r = .828$; $p < .001$) and incongruent trials ($r = .843$; $p < .001$). Bias scores were uncorrelated ($r = .007$; $p = .858$).

Discussion

The present study tested the effects of personalized and nonpersonalized ABM compared with placebo attention training delivered remotely using a mobile device to people with elevated PTSD symptoms. The present study is the largest trial of ABM to date. We demonstrated that large-scale remote administration of ABM via mobile devices and remote assessment of mechanisms and outcomes were feasible in a population reporting clinically high symptoms of PTSD. Inconsistent with our hypotheses, no group differences emerged for primary or secondary outcomes or for attention bias. However, the nonpersonalized ABM condition showed greater reductions in self-reported threat sensitivity during training compared with the placebo condition. All groups showed significant reductions in PCL-5 symptoms from pretraining to posttraining, and, inconsistent with our hypotheses, baseline attention bias scores were not differentially associated with treatment outcome in the active conditions as opposed to placebo conditions.

Our finding that ABM did not lead to greater symptom reduction compared with placebo attention training was surprising given the entirely neutral nature of our control condition. To match active and control groups on exposure to threatening stimuli, prior studies have generally used attention-control training as the control condition, in which threatening and neutral stimuli are presented and the probe replaces the neutral stimulus 50% of the time. The fact that ABM did not outperform an attention task that included *no* threat words speaks to the limited utility of ABM over and above placebo when administered via mobile devices. One possibility is that attention-control training is more effective for the treatment of PTSD than ABM, and its omission in the present study could be responsible for null group differences (Badura-Brack et al., 2015; Lazarov et al., 2019). Care should be taken, however, in interpreting similar effects for ABM and attention-control training conditions as evidence that both treatments are effective—rather, it is possible that effects of all types of attention training tasks are due to placebo effects, regression to the mean, or a combination of these factors. Although prior studies have shown that greater attention bias variability (a tendency to fluctuate between vigilance and avoidance of threat) is a mechanism associated with symptom reduction (Badura-Brack et al., 2015; Iacoviello et al., 2014; Kuckertz, Amir, et al., 2014), more recent analyses of simulated data question the validity of such measures (Kruijt, Field, & Fox, 2016). Given the controversies around this analytic approach, we chose to focus on the traditional assessment of attention bias in the present study.

We did find one significant group difference with nonpersonalized ABM showing greater reduction in self-reported threat sensitivity measured daily during the training period compared with placebo. This finding provides some evidence that ABM may have modified its purported mechanism in the present study but that these effects were too weak to carry through to the posttraining and follow-up symptom assessments. It should be noted, however, that the measure of threat sensitivity used in the present study has not undergone psychometric validation, and results should be interpreted with this caveat in mind. Furthermore, the typical ABM mechanism is attention bias measured via reaction times on the dot-probe task, and we did not find evidence for group differences in the change in attention bias. Thus, one explanation for null effects of ABM compared with placebo is that ABM failed to engage the cognitive target in our trial (MacLeod & Clarke, 2015).

However, despite showing group differences in symptom reduction, two previous trials in PTSD populations also failed to show significant group differences in change in attention bias (Badura-Brack et al., 2015). Furthermore, consistent with a recent meta-analysis of baseline bias in clinical trials (Kruijt et al., 2019), no attention bias was found at baseline, and attention bias was unrelated to PTSD symptom severity in our study. Although findings from the Kruijt et al. (2019) meta-analysis contradict a previous meta-analysis (Bar-Haim et al., 2007), the authors argued that their approach has unique strengths, including a much larger sample and the ability to circumvent publication bias by examining attention bias in clinical trials as opposed to research assessing attention bias as the primary aim. Taken together with our data, the findings call into question whether attention bias toward threat is a mechanism that can be effectively modified with ABM to reduce PTSD symptoms.

Although researchers have hypothesized that personalizing ABM stimuli (i.e., selecting stimuli that are personally relevant for each individual) should improve outcomes (Amir et al., 2016), no prior study has compared personalized to nonpersonalized ABM. Our findings indicate that personalization via selection of more highly threatening stimuli may be unlikely to improve ABM effects, at least in remote ABM studies. One possibility is that our method for selecting personalized word stimuli (via a recommender algorithm; Niles & O'Donovan, 2018), which prioritized maximizing the threat level of the selected words, was not the ideal method for personalization. Schoorl and colleagues (2014), for example, selected words on the basis of diagnostic interviews and identified words on the basis of “emotional value, distinctiveness, and times the word was repeated” (although they also failed to show beneficial effects of ABM with personalized stimuli). Thus, selecting words using constructs other than threat level could produce different results. Another possibility is that the selected words were *too* threatening, thus interfering with the ABM process. Future studies could instead use a recommender algorithm that compares training with moderately threatening words as opposed to highly threatening words to determine whether there is an optimal threat level for ABM.

All three groups showed significant reductions in PTSD symptoms; there was a mean decline of 8.5 points on the PCL-5 from baseline to posttraining. These data could indicate that either the repeated symptom assessment throughout the study or engagement in any attention-training task leads to symptom improvement. These findings also could be explained by regression to the mean given that participants were selected on the basis of high PCL-5 scores and global improvement was not shown for the STAI. Furthermore, these findings could be explained by placebo effects, which can be quite powerful (Wampold, Minami, Tierney, Baskin, & Bhati, 2005), given that participants were told in the consent process that the program may reduce anxiety and sensitivity for threat and that similar programs have been shown to be effective. Although these findings question the validity of theories that repeated practice of allocating attention away from threat reduces attentional bias and anxiety symptoms, they do suggest that a few minutes of daily engagement with a mobile app for 2 weeks could have positive effects on PTSD symptoms. Our findings highlight the importance of using active control conditions rather than waitlist control subjects in studies aiming to show the efficacy of mobile-based treatments because it appears that minimal interventions with no active component can significantly reduce symptoms.

Meta-analyses indicate that ABM is more effective when administered in the laboratory than at home (Jones & Sharpe, 2017). One hypothesis for the limited effects of ABM training at home is that participants are less engaged and sustain their attention to a lesser degree than when training is completed in the lab and thus fail to engage the cognitive mechanism purported to reduce anxiety symptoms. To assess this hypothesis, we conducted sensitivity analyses for participants receiving a greater dose of ABM and for “highly engaged” participants by examining group differences for participants who completed at least 10 sessions of training and those in the bottom 50th percentile for average reaction times. We did not find that group differences emerged for these more engaged subgroups of our sample, which questions the hypothesis that null findings for remote studies can be explained by limited engagement of attention at home. Furthermore, we did not find evidence that a baseline attention bias predicted better response to active ABM compared with placebo in terms of PCL, which is inconsistent with some prior findings (Amir et al., 2011; Kuckertz, Gildebrant, et al., 2014). It is important to note, however, that the split-half reliability of the dot-probe task, when calculated using difference scores, was very low, which calls into question the use of the dot probe as a reliable measure of attention bias. We found one significant moderation effect for the STAI (although the omnibus test was not statistically significant). Participants who showed a tendency to orient attention *away* from threat showed greater symptom reduction in the placebo group from posttraining to follow-up compared with the personalized and nonpersonalized groups. This finding may suggest that for participants who are already avoidant of threat, additional training to orient attention away from threat is iatrogenic compared with placebo training. This finding could indicate that encouraging greater avoidance of threat at early information processing stages is problematic for a subset of individuals with PTSD, although it is important to note that the finding did not replicate on the PCL-5.

Although group differences were not demonstrated between the active ABM conditions and placebo, our study was highly successful in demonstrating the potential impact of mobile-based interventions for PTSD in reaching a large number of people quickly with minimal resources. Participants were recruited from 48 out of 50 states, were ethnically and socioeconomically diverse, and came from cities, rural, and suburban areas. The income distribution in our sample was comparable with that of the United States in general (Fontenot, Semega, & Kollar, 2018), although individuals in lower income brackets were slightly more represented in our sample than in the United States overall (54% vs. 42% making less than \$50,000/year). The present study is also the largest study to date of ABM, surpassing the prior mobile-based ABM study by 294 participants (70% more) included in the analysis. Personnel primarily involved in data collection included one research coordinator and one postdoctoral fellow, and data collection was completed within 10 months. Furthermore, retention was high; approximately 75% of participants who engaged with the app in Session 1 completed at least 10 out of 12 total training sessions. These findings demonstrate not only the potential reach of mobile apps for improving mental health but also how remote-clinical-trial methodology could be used to increase the speed of research discoveries and could lead to more generalizable and replicable findings based on large samples.

Our study has some limitations. First, all participants completed the dot-probe assessment before training, during training, and after training to allow assessment of the purported mechanism of action of ABM. It is possible that this assessment led to improvements in the placebo group or diminished improvements in the active-treatment conditions and interfered with our ability to detect group differences. A second limitation is that we did not have an attention-control-training condition. Given that attention-control training has shown beneficial effects in PTSD and has outperformed ABM in three trials (Badura-Brack et al., 2015; Lazarov et al., 2019), the present study does not inform the growing evidence base for the efficacy of this version of the intervention. It is important to note, however, that no prior study has compared attention-control training with placebo training in PTSD, which limits our ability to determine its efficacy over a placebo. Third, although we used a multilevel model to test the effect of the intervention on attention bias, including probe location (congruent vs. incongruent) as a within-subjects factor, this analytic approach relies on calculating the difference between the two conditions to test statistical significance, which has demonstrated low reliability (McNally, Enock, Tsai, & Tousian, 2013; Schmukle, 2005; Waechter, Nelson, Wright, Hyatt, & Oakman, 2014; Waechter & Stolz, 2015). Furthermore, our analysis of split-half reliability showed similarly low reliability as has been found in prior studies (Rodebaugh et al., 2016). Thus, in our graphical depiction of the results (Fig. S2 in the Supplemental Material), we present reaction times for both congruent and incongruent trials to provide visual confirmation of our null result. Fourth, although results did not differ when we constrained our analyses to a subset of participants demonstrating faster reaction times (indicating greater attention on the task), remote administration makes it impossible to know who was fully engaged and who was not, which thereby increases heterogeneity in the sample. Fifth, it is important to note that the words we used for training were general threat-related words and not specific trauma-related words. Finally, because we did not assess psychiatric comorbidity, we were unable to control for it or examine subgroups in our analysis.

In sum, to our knowledge, the present study is the largest randomized clinical trial of ABM to date. We had sufficient statistical power to detect group differences with very small effect sizes, yet we were unable to demonstrate the superiority of ABM compared with a placebo training condition in reducing PTSD symptoms. Furthermore, we were unable to show any added benefit of personalizing the word stimuli used in ABM in comparison with nonpersonalized stimuli. Findings indicate limited utility of ABM when administered remotely in this population and demonstrate that personalization is unlikely to enhance effects. Publication of these null results is important because a number of mobile ABM apps already exist and are being distributed as effective methods to reduce anxiety. Yet we do not have evidence that ABM is effective when administered using this method. Our study does, however, demonstrate the feasibility of fast and low-resource recruitment of a large sample of people with elevated PTSD symptoms to engage in a 2-week cognitive training program. Remote-treatment trials can reach a large, diverse population, which suggests the potential impact of this method for reducing the burden of mental illness and for advancing research in behavioral interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This project was supported by the University of California, San Francisco (UCSF) Digital Mental Health Resource Allocation Grant as well as a grant from the Department of Defense Institute for Translational Neuroscience. A. N. Niles was supported by the San Francisco VA Women's Health Fellowship. A. O'Donovan was supported by National Institute of Mental Health Grant K01-MH109871 and a UCSF Hellman Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Amir N, Beard C, Burns M, & Bomyea J (2009). Attention modification program in individuals with generalized anxiety disorder. *Journal of Abnormal Psychology*, 118, 28–33. doi:10.1037/a0012589 [PubMed: 19222311]
- Amir N, Kuckertz JM, & Stregre MV (2016). A pilot study of an adaptive, idiographic, and multi-component attention bias modification program for social anxiety disorder. *Cognitive Therapy and Research*, 40, 661–671. doi:10.1007/s10608-016-9781-1 [PubMed: 27795598]
- Amir N, Taylor CT, & Donohue MC (2011). Predictors of response to an attention modification program in generalized social phobia. *Journal of Consulting and Clinical Psychology*, 79, 533–541. doi:10.1037/a0023808 [PubMed: 21707134]
- Antony MM, Bieling PJ, Cox BJ, Enns MW, & Swinson RP (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, 10, 176–181.
- Badura-Brack AS, McDermott TJ, Heinrichs-Graham E, Ryan TJ, Khanna MM, Pine DS, ... Wilson TW, (2018). Veterans with PTSD demonstrate amygdala hyper-activity while viewing threatening faces: A MEG study. *Biological Psychology*, 132, 228–232. doi:10.1016/j.biopsycho.2018.01.005 [PubMed: 29309826]
- Badura-Brack AS, Naim R, Ryan TJ, Levy O, Abend R, Khanna MM, ... Bar-Haim Y (2015). Effect of attention training on attention bias variability and PTSD symptoms: Randomized controlled trials in Israeli and U.S. combat veterans. *American Journal of Psychiatry*, 172, 1233–1241. doi:10.1176/appi.ajp.2015.14121578
- Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, & van IJzendoorn MH (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, 133, 1–24. [PubMed: 17201568]
- Beck JG, Freeman JB, Shipperd JC, Hamblen JL, & Lackner JM (2001). Specificity of Stroop interference in patients with pain and PTSD. *Journal of Abnormal Psychology*, 110, 536–543. [PubMed: 11727943]
- Blevins CA, Weathers FW, Davis MT, Witte TK, & Domino JL (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, 28, 489–498. doi:10.1002/jts.22059 [PubMed: 26606250]
- Blix I, & Brennen T (2011). Intentional forgetting of emotional words after trauma: A study with victims of sexual assault. *Frontiers in Psychology*, 2, Article 235. doi:10.3389/fpsyg.2011.00235
- Boettcher J, Berger T, & Renneberg B (2012). Internet-based attention training for social anxiety: A randomized controlled trial. *Cognitive Therapy and Research*, 3, 522–536. doi:10.1007/s10608-011-9374-y
- Bradley MM, & Lang PJ (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings* (Technical Report C-1). Center for Research in Psychophysiology, University of Florida, Gainesville.

- Bunnell SL (2007). The impact of abuse exposure on memory processes and attentional biases in a college-aged sample (Master's thesis, University of Kansas). Retrieved from <https://pdfs.semanticscholar.org/9075/9ec30d17546e5038f8d4d1a63332d77ffef2.pdf>
- Carlbring P, Apelstrand M, Sehlin H, Amir N, Rousseau A, Hofmann SG, & Andersson G (2012). Internet-delivered attention bias modification training in individuals with social anxiety disorder—a double blind randomized controlled trial. *BMC Psychiatry*, 12, Article 66. doi:10.1186/1471-244X-12-66
- Carlson EB, Smith SR, Palmieri PA, Dalenberg C, Ruzek JI, Kimerling R, ... Spain DA (2011). Development and validation of a brief self-report measure of trauma exposure: The trauma history screen. *Psychological Assessment*, 23, 463–477. doi:10.1037/a0022294 [PubMed: 21517189]
- Craske MG, Niles AN, Burklund LJ, Wolitzky-Taylor KB, Plumb C, Arch JJ, ... Lieberman MD (2014). Randomized controlled trial of cognitive behavioral therapy and acceptance and commitment therapy for social phobia: Outcomes and moderators. *Journal of Consulting and Clinical Psychology*, 82, 1034–1048. doi:10.1037/a0037212 [PubMed: 24999670]
- Cristea IA, Kok RN, & Cuijpers P (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *The British Journal of Psychiatry: The Journal of Mental Science*, 206, 7–16. doi:10.1192/bjp.bp.114.146761 [PubMed: 25561486]
- Cusack K, Jonas DE, Forneris CA, Wines C, Sonis J, Middleton JC, ... Gaynes BN (2016). Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. *Clinical Psychology Review*, 43, 128–141. doi:10.1016/j.cpr.2015.10.003 [PubMed: 26574151]
- Davies M (2008) The Corpus of Contemporary American English (COCA): 600 million words, 1990-present. Available at <https://www.english-corpora.org/coca/>
- Derryberry D, & Reed MA (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology*, 111, 225–236. [PubMed: 12003445]
- Ehring T, & Ehlers A (2011). Enhanced priming for trauma-related words predicts posttraumatic stress disorder. *Journal of Abnormal Psychology*, 120, 234–239. doi:10.1037/a0021080 [PubMed: 21058753]
- Enock PM, Hofmann SG, & McNally RJ (2014). Attention bias modification training via smartphone to reduce social anxiety: A randomized, controlled multi-session experiment. *Cognitive Therapy and Research*, 38, 200–216. doi:10.1007/s10608-014-9606-z
- Fani N, Tone EB, Phifer J, Norrholm SD, Bradley B, Ressler KJ, ... Jovanovic T (2012). Attention bias toward threat is associated with exaggerated fear expression and impaired extinction in PTSD. *Psychological Medicine*, 42, 533–543. doi:10.1017/S0033291711001565 [PubMed: 21854700]
- Faul F, Erdfelder E, Lang A-G, & Buchner A (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. [PubMed: 17695343]
- Feingold A (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, 14, 43–53. doi:10.1037/a0014699 [PubMed: 19271847]
- Fontenot K, Semega J, & Kollar M (2018). Income and poverty in the United States: 2017 (Report No. P60–26). Retrieved from the U.S. Census Bureau website: <https://www.census.gov/library/publications/2018/demo/p60-263.html>
- Iacoviello BM, Wu G, Abend R, Murrrough JW, Feder A, Fruchter E, ... Charney DS (2014). Attention bias variability and symptoms of posttraumatic stress disorder. *Journal of Traumatic Stress*, 27, 232–239. doi:10.1002/jts.21899 [PubMed: 24604631]
- Jones EB, & Sharpe L (2017). Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders*, 223, 175–183. doi:10.1016/j.jad.2017.07.034 [PubMed: 28759865]
- Kessler RC, Aguilar-Gaxiola S, Alonso J, Benjet C, Bromet EJ, Cardoso G, ... Koenen KC (2017). Trauma and PTSD in the WHO World Mental Health Surveys. *European Journal of Psychotraumatology*, 8(Suppl. 5), Article 1353383. doi:10.1080/20008198.2017.1353383
- Khanna MM, Badura-Brack AS, McDermott TJ, Shepherd A, Heinrichs-Graham E, Pine DS, ... Wilson TW (2016). Attention training normalises combat-related post-traumatic stress disorder

- effects on emotional Stroop performance using lexically matched word lists. *Cognition & Emotion*, 30, 1521–1528. doi:10.1080/02699931.2015.1076769 [PubMed: 26309165]
- Kousta S-T, Vinson DP, & Vigliocco G (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112, 473–481. doi:10.1016/j.cognition.2009.06.007 [PubMed: 19591976]
- Kruijt A-W, Field AP, & Fox E (2016). Capturing dynamics of biased attention: Are new attention variability measures the way forward? *PLOS ONE*, 11(11), Article e0166600. doi:10.1371/journal.pone.0166600
- Kruijt A-W, Parsons S, & Fox E (2019). A meta-analysis of bias at baseline in RCTs of attention bias modification: No evidence for attention bias towards threat in clinical anxiety and PTSD. *Abnormal Psychology*, 128, 563–573. doi:10.1037/abn0000406
- Kruijt A-W, Putman P, & Van der Does W (2013). The effects of a visual search attentional bias modification paradigm on attentional bias in dysphoric individuals. *Journal of Behavior Therapy and Experimental Psychiatry*, 44, 248–254. doi:10.1016/j.jbtep.2012.11.003 [PubMed: 23266603]
- Kuckertz JM, Amir N, Boffa JW, Warren CK, Rindt SEM, Norman S, ... McLay R (2014). The effectiveness of an attention bias modification program as an adjunctive treatment for post-traumatic stress disorder. *Behaviour Research and Therapy*, 63, 25–35. doi:10.1016/j.brat.2014.09.002 [PubMed: 25277496]
- Kuckertz JM, Gildebrant E, Liliequist B, Karlström P, Våppling C, Bodlund O, ... Carlbring P (2014). Moderation and mediation of the effect of attention training in social anxiety disorder. *Behaviour Research and Therapy*, 53, 30–40. doi:10.1016/j.brat.2013.12.003 [PubMed: 24373984]
- Lazarov A, Suarez-Jimenez B, Abend R, Naim R, Shvil E, Helpman L, ... Neria Y (2019). Bias-contingent attention bias modification and attention control training in treatment of PTSD: A randomized control trial. *Psychological Medicine*, 49, 2432–2440. doi:10.1017/S0033291718003367 [PubMed: 30415648]
- Linetsky M, Pergamin-Hight L, Pine DS, & Bar-Haim Y (2015). Quantitative evaluation of the clinical efficacy of attention bias modification treatment for anxiety disorders. *Depression and Anxiety*, 32, 383–391. doi:10.1002/da.22344 [PubMed: 25708991]
- MacLeod C, & Clarke PJF (2015). The attentional bias modification approach to anxiety intervention. *Clinical Psychological Science*, 3, 58–78. doi:10.1177/2167702614560749
- McNally RJ, Enock PM, Tsai C, & Tousian M (2013). Attention bias modification for reducing speech anxiety. *Behaviour Research and Therapy*, 51, 882–888. doi:10.1016/j.brat.2013.10.001 [PubMed: 24211339]
- Najmi S, & Amir N (2010). The effect of attention training on a behavioral test of contamination fears in individuals with subclinical obsessive-compulsive symptoms. *Journal of Abnormal Psychology*, 119, 136–142. doi:10.1037/a0017549 [PubMed: 20141250]
- National Center for PTSD. (2018, 9). PTSD Checklist for DSM-5 (PCL-5). Retrieved from <https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp>
- Neubauer K, von Auer M, Murray E, Petermann F, Helbig-Lang S, & Gerlach AL (2013). Internet-delivered attention modification training as a treatment for social phobia: A randomized controlled trial. *Behaviour Research and Therapy*, 51, 87–97. doi:10.1016/j.brat.2012.10.006 [PubMed: 23262116]
- Niles AN, Craske MG, Lieberman MD, & Hur C (2015). Affect labeling enhances exposure effectiveness for public speaking anxiety. *Behaviour Research and Therapy*, 68, 27–36. doi:10.1016/j.brat.2015.03.004 [PubMed: 25795524]
- Niles AN, Luxenberg A, Neylan TC, Inslicht SS, Richards A, Metzler TJ, ... O'Donovan A (2018). Effects of threat context, trauma history, and posttraumatic stress disorder status on physiological startle reactivity in Gulf War veterans. *Journal of Traumatic Stress*, 31, 579–590. doi:10.1002/jts.22302 [PubMed: 30058728]
- Niles AN, & O'Donovan A (2018). Personalizing affective stimuli using a recommender algorithm: An example with threatening words for trauma exposed populations. *Cognitive Therapy and Research*, 42, 747–757. doi:10.1007/s10608-018-9923-8 [PubMed: 31571703]
- Pew Research Center. (2019612). Mobile fact sheet. Retrieved from <http://www.pewinternet.org/fact-sheet/mobile/>

- Powers MB, Halpern JM, Ferenschak MP, Gillihan SJ, & Foa EB (2010). A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clinical Psychology Review*, 30, 635–641. doi:10.1016/j.cpr.2010.04.007 [PubMed: 20546985]
- Price RB, Kuckertz JM, Siegle GJ, Ladouceur CD, Silk JS, Ryan ND, ... Amir N (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment*, 27, 365–376. doi:10.1037/pas0000036 [PubMed: 25419646]
- Roberts KE, Hart TA, & Eastwood JD (2010). Attentional biases to social and health threat words in individuals with and without high social anxiety or depression. *Cognitive Therapy and Research*, 34, 388–399. doi:10.1007/s10608-009-9245-y
- Rodebaugh TL, Scullin RB, Langer JK, Dixon DJ, Huppert JD, Bernstein A, ... Lenze EJ (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of Abnormal Psychology*, 125, 840–851. doi:10.1037/abn0000184 [PubMed: 27322741]
- Roy-Byrne PP, Craske MG, Sullivan G, Rose RD, Edlund MJ, Lang AJ, ... Stein MB (2010). Delivery of evidence-based treatment for multiple anxiety disorders in primary care. *JAMA*, 303, 1921–1928. doi:10.1001/jama.2010.608 [PubMed: 20483968]
- Schmukle SC (2005). Unreliability of the dot probe task. *European Journal of Personality*, 19, 595–605. doi:10.1002/per.554
- Schoorl M, Putman P, Mooren TM, Van Der Werff S, & Van Der Does W (2014). Attentional bias modification in Dutch veterans with posttraumatic stress disorder: A case series with a personalized treatment version. *Journal of Traumatic Stress*, 27, 240–243. doi:10.1002/jts.21896 [PubMed: 24700603]
- Schoorl M, Putman P, & Van Der Does W (2013). Attentional bias modification in posttraumatic stress disorder: A randomized controlled trial. *Psychotherapy and Psychosomatics*, 82, 99–105. doi:10.1159/000341920 [PubMed: 23295710]
- Spielberger CD (1985). Assessment of state and trait anxiety: Conceptual and methodological issues. *Southern Psychologist*, 2(4), 6–16.
- Swick D (2009). The separate and cumulative effects of TBI and PTSD on cognitive function and emotional control (Annual Report No. W81XWH-08-2-0086). Martinez, CA: Veterans Health Administration.
- Thompson SC, Schlehofer MM, & Bovin MJ (2006). The measurement of threat orientations. *American Journal of Health Behavior*, 30, 147–157. doi:10.5555/ajhb.2006.30.2.147 [PubMed: 16533099]
- Vasterling JJ, & Arditte Hall KA (2018). Neurocognitive and information processing biases in posttraumatic stress disorder. *Current Psychiatry Reports*, 20(11), Article 99. doi:10.1007/s11920-018-0964-1
- Waechter S, Nelson AL, Wright C, Hyatt A, & Oakman J (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research*, 38, 313–333. doi:10.1007/s10608-013-9588-2
- Waechter S, & Stolz JA (2015). Trait anxiety, state anxiety, and attentional bias to threat: Assessing the psychometric properties of response time measures. *Cognitive Therapy and Research*, 39, 441–458. doi:10.1007/s10608-015-9670-z
- Wampold BE, Minami T, Tierney SC, Baskin TW, & Bhati KS (2005). The placebo is powerful: Estimating placebo effects in medicine and psychotherapy from randomized clinical trials. *Journal of Clinical Psychology*, 61, 835–854. doi:10.1002/jclp.20129 [PubMed: 15827993]
- Wolitzky-Taylor K, Niles AN, Ries R, Krull JL, Rawson R, Roy-Byrne P, & Craske M (2018). Who needs more than standard care? Treatment moderators in a randomized clinical trial comparing addiction treatment alone to addiction treatment plus anxiety disorder treatment for comorbid anxiety and substance use disorders. *Behaviour Research and Therapy*, 107, 1–9. doi:10.1016/j.brat.2018.05.005 [PubMed: 29800622]
- Woud ML, Verwoerd J, & Krans J (2017). Modification of cognitive biases related to posttraumatic stress: A systematic review and research agenda. *Clinical Psychology Review*, 54, 81–95. doi:10.1016/j.cpr.2017.04.003 [PubMed: 28445840]

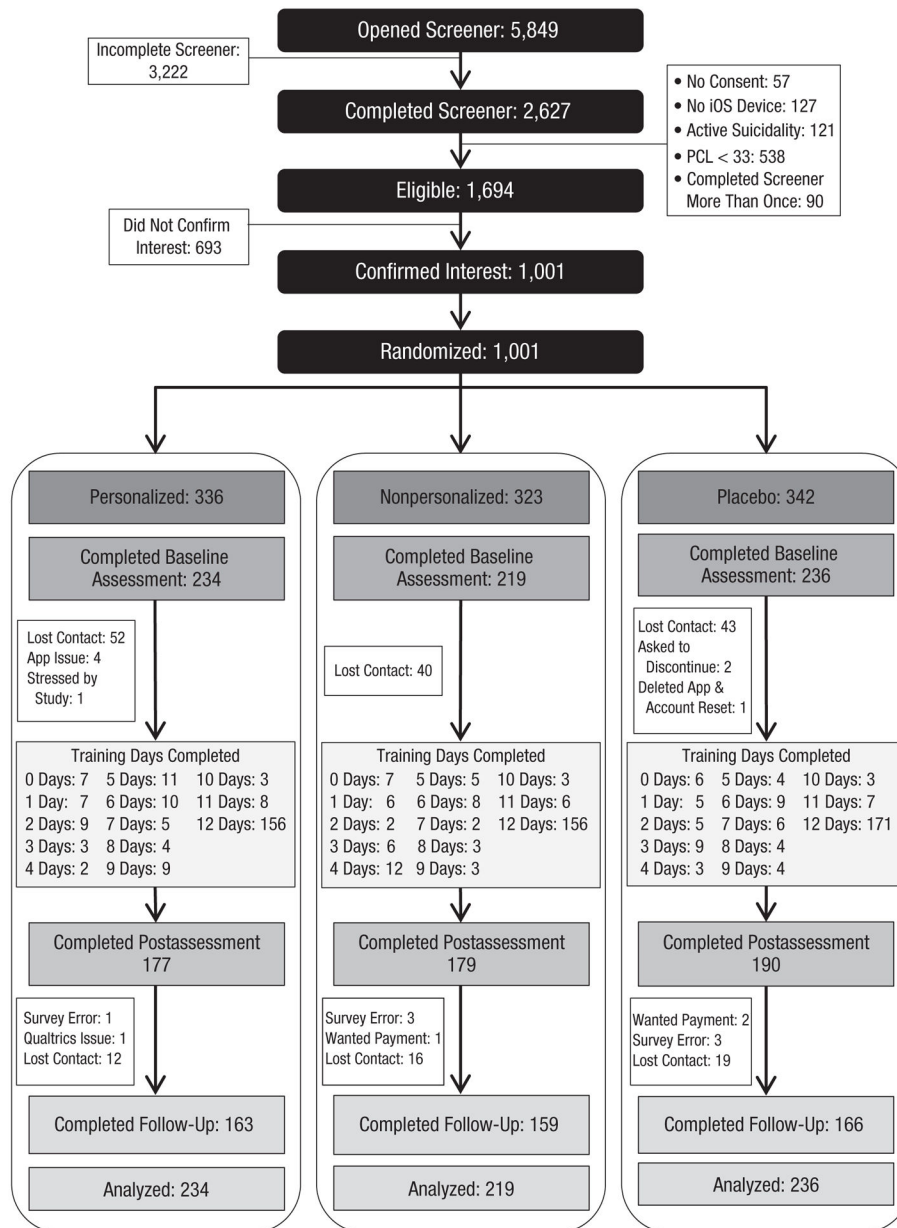


Fig. 1. Consolidated Standards of Reporting Trials (CONSORT) diagram of participant flow through the study.

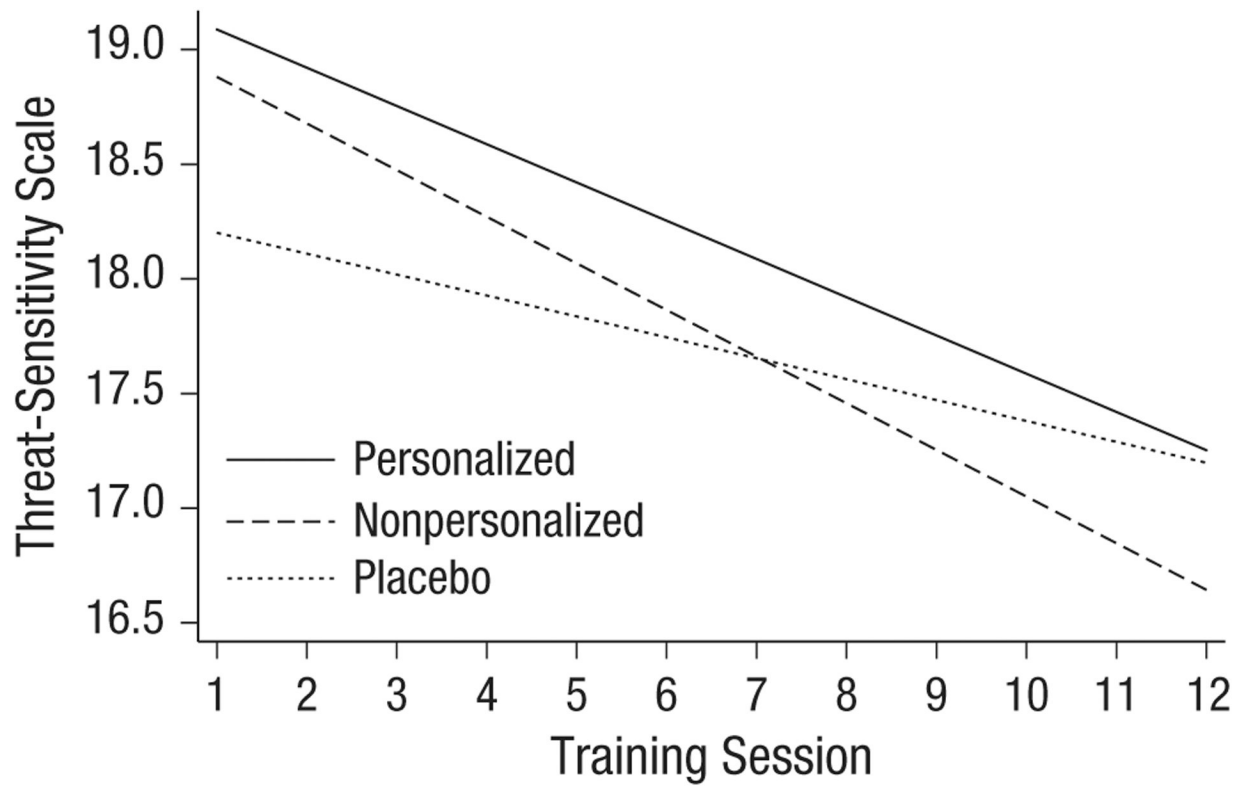


Fig. 2.
Effect of group on self-reported threat sensitivity during training.

Table 1.

Demographic and Clinical Characteristics of the Analyzed Sample

Characteristic	Value
Female	551 (80.0)
Age	32.1 (9.9)
Race	
Hispanic	89 (12.9)
Asian	57 (8.3)
Black	85 (12.4)
White	399 (60.0)
Other/more than one race	58 (8.4)
Education	
High school or less	90 (13.1)
Some college/associate's degree	247 (35.9)
4-year college or graduate school	351 (51.0)
Employed	
Unemployed	174 (25.3)
Employed part-time	115 (16.7)
Employed full-time	303 (44.0)
Student	96 (13.4)
Income	
< \$50,000	375 (54.5)
\$50,000–\$100,000	222 (32.3)
\$100,001–\$150,000	64 (9.3)
> \$150,000	27 (3.9)
Marital status	
Married	189 (27.5)
Single, in committed relationship	244 (35.5)
Single, no relationship	197 (28.6)
Separated/divorced/widowed	58 (8.4)
Location type	
City	363 (53.3)
Suburb	205 (30.1)
Town	69 (10.1)
Rural	44 (6.5)
Current psychiatric medications	256 (37.2)
Depression	182 (71.1)
Anxiety	194 (75.8)
Serious mental illness	37 (14.5)
Psychotherapy (within 6 months)	358 (52.1)
Cognitive behavioral therapy	83 (23.2)
Measures	

Characteristic	Value
PTSD Checklist, mean (<i>SD</i>)	45.7 (14.5)
Intrusion, mean (<i>SD</i>)	10.9 (4.3)
Avoidance, mean (<i>SD</i>)	4.7 (2.1)
Negative cognitions/mood, mean (<i>SD</i>)	17.1 (6.0)
Arousal/reactivity, mean (<i>SD</i>)	12.9 (5.0)
State Trait Anxiety Inventory, mean (<i>SD</i>)	48.3 (10.2)
DASS – Depression subscale, mean (<i>SD</i>)	10.6 (5.5)

Note: Values are *zs* with percentages in parentheses unless noted otherwise. PTSD = posttraumatic stress disorder; DASS = Depression Anxiety and Stress Scale

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2. Results From Multilevel Models for the Time × Condition Interaction Effect on Primary and Secondary Symptom Outcomes

	Personalized ABM versus nonpersonalized ABM				Nonpersonalized ABM versus placebo				Personalized ABM versus placebo									
	Baseline to posttraining		Posttraining to follow-up		Baseline to posttraining		Posttraining to follow-up		Baseline to posttraining		Posttraining to follow-up							
	<i>b</i>	95% CI	<i>d</i>	<i>b</i>	95% CI	<i>d</i>	<i>b</i>	95% CI	<i>d</i>	<i>b</i>	95% CI	<i>d</i>						
Primary outcomes																		
PTSD Checklist	0.6	[-2.5, 3.7]	0.0	1.1	[-2.2, 4.4]	0.1	-1.0	[-4.1, 2.1]	0.1	-1.0	[-4.3, 2.3]	-0.1	-0.4	[-3.5, 2.6]	0.0	0.1	[-3.2, 3.3]	0.0
State Trait Anxiety Inventory	1.1	[-0.5, 2.6]	-0.1	-0.4	[-2.0, 1.2]	0.0	-1.3	[-2.9, 0.2]	0.1	1.2	[-0.4, 2.8]	0.1	-0.3	[-1.8, 1.2]	0.0	0.8	[-0.8, 2.4]	0.1
Secondary outcomes																		
PTSD Checklist subscales																		
Intrusion	0.4	[-0.5, 1.4]	-0.1	0.1	[-0.9, 1.1]	0.0	-0.2	[-1.1, 0.7]	0.1	0.0	[-1.0, 1.0]	0.0	0.2	[-0.7, 1.1]	0.0	0.1	[-0.9, 1.1]	0.0
Avoidance	-0.2	[-0.7, 0.3]	-0.1	0.0	[-0.5, 0.5]	0.0	0.3	[-0.2, 0.8]	0.1	-0.2	[-0.7, 0.3]	0.1	0.1	[-0.3, 0.6]	0.1	-0.2	[-0.7, 0.3]	-0.1
Negative cognitions/ mood	0.0	[-1.3, 1.3]	0.0	0.7	[-0.6, 2.1]	0.1	0.4	[-0.9, 1.7]	0.1	-0.3	[-1.6, 1.1]	0.0	0.4	[-0.9, 1.7]	0.1	0.5	[-0.9, 1.8]	0.1
Arousal/reactivity	0.0	[-1.0, 1.1]	0.0	0.2	[-0.9, 1.3]	0.0	0.0	[-1.0, 1.1]	0.0	-0.5	[-1.6, 0.6]	-0.1	0.1	[-0.9, 1.1]	0.0	-0.3	[-1.4, 0.8]	-0.1
DASS-Depression subscale	-0.5	[-1.5, 0.4]	-0.1	0.2	[-0.8, 1.1]	0.0	0.3	[-0.6, 1.2]	0.1	0.3	[-0.7, 1.2]	0.0	-0.2	[-1.2, 0.7]	0.0	0.4	[-0.6, 1.4]	0.1

Note: ABM = attention-bias modification; PTSD = posttraumatic stress disorder; DASS = Depression Anxiety and Stress Scale; CI = confidence interval; *b* = unstandardized regression coefficient; *d* = Cohen's *d* (Feingold, 2009).