Fear conditioning and extinction learning in the mood and anxiety disorders spectrum – Associations with the outcome of cognitive behavior therapy

Dirk Adolph *, Tobias Teismann, Andre Wannemüller, Jürgen Margraf

Mental Health Research and Treatment Center, Ruhr University Bochum, Germany

ARTICLE INFO

Keywords:
Fear extinction
Safety learning
Depression
Anxiety
Cognitive behavior therapy
Prediction of treatment outcome

ABSTRACT

In the current study, we test for the specificity of deficits in fear acquisition and extinction for the anxiety disorders spectrum. We compared fear acquisition and fear extinction learning between a group of patients with either an anxiety disorder (n = 93) or depression (n = 103) attending for treatment in our outpatient center and a sample of healthy control participants (n = 60). To assess the specificity of the predictive validity of extinction learning and safety learning for the outcome of exposure-based treatments, patients additionally underwent disorder-specific cognitive behavior therapy (CBT). We found only very little evidence for differences in safety or extinction learning between healthy controls and patients with anxiety-disorders or depression using both a group-based categorical analytic approach, as well as a trans-diagnostic, dimensional analytic approach. On the contrary, for anxiety patients only, more favorable extinction learning and more favorable safety learning was associated with more favorable treatment outcome. In sum, this specific prediction of treatment outcome in anxiety patients confirms and extends current theoretical models of exposure-based treatments for anxiety disorders, but does not support the notion of general extinction learning deficits in the anxiety disorders spectrum.

1. Introduction

Fear conditioning and fear extinction learning constitute central mechanisms within etiological models of anxiety disorders (for an overview see Bouton, Mineka, & Barlow, 2001; Susan Mineka & Zinbarg, 2006) and models explaining the reduction of pathological fear during exposure-based treatments (In-Albon & Schneider, 2007; McGuire, Lewin, & Storch, 2014; Milad & Quirk, 2012; Vervliet, Craske, & Hermans, 2013). Supporting these theoretical considerations, meta-analyses have shown that patients with anxiety disorders in contrast to healthy controls show enhanced fear responses towards conditioned safety cues during fear acquisition (i.e., the conditioned stimulus never paired with the aversive consequence, CS−) (Duits et al., 2015; Lissek et al., 2005) and exaggerated fear responses towards conditioned fear cues during extinction learning (i.e., the conditioned stimulus previously paired with the aversive consequence, CS+). In addition, initial evidence suggests predictive validity of extinction learning (see Scheveneels, Boddez, Vervliet, & Hermans, 2016) for the outcome of exposure-based treatments in the anxiety disorders spectrum (review in Scheveneels, Boddez, & Hermans, 2021). However, to date, only very little research has been conducted on the specificity of these findings with respect to the anxiety disorders spectrum. Within the present study, we aimed at closing this gap and compared fear acquisition and fear extinction learning between a large group of patients with either an anxiety or depressive disorder and a sample of healthy controls. To test for the specificity of the predictive validity of extinction learning for the outcome of exposure-based treatments, patients additionally underwent disorder-specific cognitive behavior therapy (CBT).

Indeed, there is still considerable debate on how affective and anxiety disorders might be separable in terms of their underlying pathogenic processes. Both disorders are highly comorbid (Gorman, 1996; Kessler, Chiu, Demler, Merikangas, & Walters, 2005; Sartorius, Ustian, Lecrubier, & Wittchen, 1996) and share a considerable amount of symptomatology (for an overview see Clark & Watson, 1991). Hence, factor-analytical work has repeatedly found depression and anxiety to be characterized by the same underlying psychopathological processes, best described as an internalizing dimension (Achenbach, 1966; Krueger, 1999; Krueger & Markon, 2006), indicating a liability to experience enhanced negative affectivity (i.e., mood and anxiety symptoms, for an overview of current models see Shankman & Klein)

* Corresponding author. Massenbergerstraße 9-13, D-44787, Bochum, Germany.
E-mail address: dirk.adolph@rub.de (D. Adolph).

https://doi.org/10.1016/j.brat.2022.104229
Received 7 February 2022; Received in revised form 8 November 2022; Accepted 11 November 2022
Available online 17 November 2022
0005-7967/© 2022 Elsevier Ltd. All rights reserved.
However, despite these broad similarities, there is consensus between current models (for an overview of these models see Shankman & Klein, 2003) that many anxiety disorders (for example panic disorder, specific phobias or PTSD), but not depression, are related to elevated threat responses and physiological hyperarousal (Dillon et al., 2014; Joiner et al., 1999; McTeague & Lang, 2012; Shankman & Klein, 2003). In turn, with some exceptions (e.g. social anxiety disorder) low positive affect is mainly related to depression rather than anxiety disorders (e.g. Shankman & Klein, 2003). Thus, there is considerable evidence showing that patients suffering from depression and anxiety are also distinguishable on basis of their affective reactivity (Mineka, Watson, & Clark, 1998; Shankman & Klein, 2003).

Within this framework, fear conditioning and extinction learning are promising mechanisms to differentiate between depression and anxiety. That is, supporting differences in affective processing, and in direct contrast to anxiety, research so far has shown that patients suffering from depression show enhanced extinction learning as compared to healthy controls (Kuhn et al., 2014), as well as better differential acquisition of fear responses towards the CS+ (Nissen et al., 2010), probably mediated by less pronounced deficits in safety learning (i.e., less pronounced fear responses toward the CS– see Jovanovic et al., 2010). Notably, there are also reports showing contrary (Otto et al., 2014), or null findings (Dibbets, van den Broek, & Evers, 2015). Furthermore, previous studies mostly rely on comparably small sample sizes (i.e. ranging from n = 13 to n = 37 patients per diagnostic group). Thus, further research is needed to clarify common and distinct effects of anxiety and depression on fear acquisition and extinction learning.

In addition, a direct comparison of patients with depression and anxiety is ideally suited to test for the specificity of extinction learning as an anxiety-specific CBT mechanism (i.e., concerning exposure-based treatment approaches). Because anxiety and depression exhibit a large amount of common but also distinct symptomatology and pathological processes (i.e., for a comprehensive discussion see Clark & Watson, 1991), CBT approaches for both disorders share common features, like cognitive restructuring or contingency management but are also distinguishable in terms of exposure techniques which are provided almost exclusively during the treatment for anxiety disorders but not depression (for a comprehensive overview of CBT treatment protocols see Hofmann, Dozois, Rief, & Smits, 2014). Thus, if extinction learning is a valid laboratory mechanism for symptom reduction during anxiety treatment via exposure techniques, extinction learning in the laboratory should be predictive of treatment outcome in CBT for anxiety disorders, but not depression. So far, only one study directly addressed the specificity of the predictive validity of extinction learning for exposure-based treatments. Lange et al. (2020) compared the predictive validity of extinction learning for a one-session exposure-based treatment and a one-session relaxation approach in a sample of adult patients with spider phobia. They found that neural responses during pre-treatment extinction learning was associated with the outcome of the exposure-based treatment, but not with the outcome of the relaxation intervention. No study has directly compared the predictive validity of extinction learning for the outcome of CBT approaches to anxiety and depression.

1.1. The current study

Taken together, within the current study, we assessed fear conditioning and extinction learning in healthy controls and patients with either anxiety or depression prior to a disorder-specific CBT treatment. Importantly, we did not exclude patients with comorbid diagnoses, because we aimed for ecologically valid patient-samples who are representative for the patients typically receiving treatment in our outpatient center.

Based on the current literature outlined above, we awaited deficient extinction learning (i.e., elevated responses towards the CS+ during extinction) and deficient safety learning (i.e., elevated responses towards the CS–) in patients with anxiety disorders, as compared to depressed patients, and healthy controls. Moreover, we awaited the outcome of CBT to be predicted by extinction learning in patients with anxiety disorders, but not in patients with depression. In addition, specifically emphasizing safety learning, we also explored the association between markers of fear acquisition and treatment outcome. However, because previous research yield contradictory findings, showing that either superior (Duits et al., 2021) or maladaptive (Wannemüller et al., 2018) safety learning predicts more favorable treatment outcome, no specific hypotheses could be drawn from the literature.

Finally, in addition to categorical diagnosis-based analyses, we implemented a dimensional, transdiagnostic approach to our data. Traditional classification approaches of mental disorders show only limited diagnostic reliability. They mostly fail to account for frequently occurring diagnostic comorbidity between disorders (Kessler et al., 2005) and symptom-based heterogeneity within disorder subgroups (Widiger & Trull, 2007). It has been suggested that these limitations are direct consequence of categorizing dimensional phenomena (Forbes et al., 2021). Thus, in addition to the comparison of indicators of fear acquisition and extinction learning between diagnostic groups, we aimed for a transdiagnostic approach which is broadly compatible with the ideas from the NIMH’s Research Domain Criteria (RDoC) initiative. That is, rather than solely constituting qualitatively uniform entities, we also consider mental disorders to fall along a continuum ranging from normal to disordered functioning (Cuthbert, 2014). In analyzing data in a transdiagnostic manner (i.e., across current diagnostic categories and healthy controls), this approach potentially adds to our understanding of current concepts of diagnosis (Cuthbert, 2014). In addition, it may help to identify new disorder subgroups, and to define borders between normal and pathological functioning.

2. Methods and materials

2.1. Sample characteristics

A total of N = 223 patients and N = 60 healthy controls (HC) agreed to participate in the current study. All participants gave written informed consent to procedures. The study was conducted in accord with the Declaration of Helsinki and was approved by the ethics review board of the Faculty of Psychology at Ruhr-University Bochum.

Healthy controls were recruited from the community via advertisements. Patients were recruited from the outpatient clinic of the Mental Health Research and Treatment Center (MHRTC) at Ruhr University Bochum. All patients attended for treatment at MHRTC and were diagnosed during a regular diagnostic session prior to the treatment. Structured interviews were performed by trained postgraduate psychotherapists using a standardized interview for DSM-IV disorders (DIPS, Schneider & Margraf, 2006). Assessment of mental disorders within the control group was done using a brief semi-structured interview for DSM-IV disorders (Mini-Dips, Margraf, 1994) and via self-report. No control participant had to be excluded due to a current or history of mental disorders.

Of the initial N = 223 patients, n = 27 did not qualify for a diagnosis of a depressive or an anxiety disorder and had to be excluded from the current study. There were no further exclusion criteria. Of the remaining 196 patients, n = 103 patients qualified for a diagnosis of a depressive disorder (n = 91 Major Depression, n = 9 Dysthymia, n = 3 other depressive disorder) and n = 93 patients were diagnosed for an anxiety disorder (n = 28 social phobia; n = 25 panic disorder/agoraphobia, n =
14 generalized anxiety disorder, n = 8 obsessive compulsive disorder, n = 8 post traumatic stress disorder, n = 6 specific phobia, n = 4 other anxiety disorder). In addition, a total of n = 22 patients with depression (DEP) had a comorbid anxiety disorder and a total of n = 31 patients with an anxiety disorder (ANX) had a comorbid depressive disorder. Patients with comorbid anxiety and depression were grouped on basis of their primary diagnosis (i.e. the disorder they attended to be treated for).患者可能在其它障碍（即，符合障碍）受影响。比例（i.e. 在此病例中为抑郁和焦虑障碍）。

Five participants discontinued the experiment without giving a reason (n = 2 DEP, n = 1 ANX, n = 2 HC) so that n = 251 participants finished the conditioning/extinction task (n = 58 HC, n = 92 ANX, n = 101 DEP). Of these, n = 95 were male (HC n = 14, DEP n = 42, ANX n = 39) and n = 156 were female (HC n = 44, DEP n = 59, ANX n = 53). Results indicate that there were more female participants in the group of HC as compared to the anxiety, \( \chi^2 (1) = 5.57, p = .018 \), and the depression group, \( \chi^2 (1) = 5.12, p = .024 \). The final sample was between 18 and 73 years of age (HC = 18–59, M = 33.19, SD = 11.29; DEP = 18–69, M = 39.54, SD = 14.68; ANX = 19–73, M = 34.14, SD = 11.91). Depressed patients were slightly older than anxiety patients, p = .004, and healthy controls, p = .003, F(2,249) = 6.04, p = .003.

2.2. Fear conditioning and extinction task

The conditioning and extinction task closely resemble a task used previously in clinical populations (Blechert, Michael, Friends, Margraf, & Wilhelm, 2007; Michael, Blechert, Friends, Margraf, & Wilhelm, 2007). It consists of a pre-acquisition, an acquisition and an extinction phase. The three phases were presented consecutively without a break. Throughout the phases two Rorschach inkblots served as CSs (i.e., CS+, CS-) and were presented in random order (with the exception that each CS was presented no more than two times in succession) for 7 s with an ISI varying between 16 and 18 s. During the pre-acquisition phase, each CS was presented three times, whereas during the acquisition and the extinction phases each CS was presented six times. During acquisition, only one of the CSs (i.e., the CS+) was immediately followed by a mild electrical shock (i.e., the unconditioned stimulus, US) presented for 500 ms (reinforcement rate = 100%). Across participants, each of the two inkblots served as CS+ and CS- equally often.

Prior to the beginning of the task, participants were told that they will see several pictures and that one picture might sometimes be followed by a mild electrical shock. After that, participants individually adjusted the shock as described previously (Klampers et al., 2010). They then rated the shock for unpleasantness (i.e., US rating), 0 = not at all unpleasant, 100 = extremely unpleasant). Thereafter the task began. After each phase, participants rated the CSs for valence (i.e., 0 = absolutely sure, see (Blechert et al., 2007 for details).

2.3. Treatment

The patients received typical treatments as routinely provided in our outpatient center. All treatments were paid for by the German health care insurance system and included approximately 25 sessions of Cognitive Behavior Therapy (CBT). The number of sessions did not differ between patients with depressive or anxiety disorders (DEP, M = 25.4, SD = 3.7; ANX, M = 25.9, SD = 2.6, p = .382). All treatments were provided by therapists in postgraduate training at the outpatient clinic of the Mental Health Research and Treatment Center at Ruhr University Bochum, Germany. Therapists at MHRTC have at least a Master’s degree in Psychology and a minimum of 1-year full-time postgraduate training in CBT. Generally, all treatments at MHRTC follow CBT manuals for the respective disorders (i.e., in this case for depressive and anxiety disorder). A comprehensive list of manuals used for anxiety disorder treatments are given in the supplemental materials. Importantly, all of these manuals are exposure-based. As part of postgraduate training, treatments are additionally monitored by licensed CBT supervisors. During these regular supervision sessions, therapists and the supervisor discuss the patient’s current status and the ongoing treatment, including the treatment plan. However, despite general agreement with published manuals (i.e., exposure-based CBT for anxiety disorders) and ongoing supervision, treatments within routine outpatient care are usually less standardized than in typical randomized controlled trials (von Brachel et al., 2019).

2.4. Physiological data assessment and data reduction

Electrodermal activity was assessed continuously at 1000Hz sampling rate using a Biopac MP100 amplifier system. Offline, electrodermal activity data was highpass-filtered (1Hz, 24dB/occt). First interval skin conductance responses (FIR) was then quantified as the difference between the foot-point occurring between 1 and 4 s after stimulus onset and the consecutive maximum deflection of the current response. Response quantification was done according to published guidelines (Boucsein, 2012). Trials with no detectable responses were scored as zero. SCRs were then square root transformed and mean SCR were calculated for the CS+ and the CS- within each of the three experimental phases.

2.5. Assessment of treatment outcome

Treatment outcome was monitored via questionnaire-based pre-post symptom assessment using the brief version of the Depression, Anxiety and Stress Scale (DASS-21, Lovibond & Lovibond, 1995). The validity of the DASS-21 for the use in clinical populations (Bottesi et al., 2015) and as a treatment outcome measure has been demonstrated (Ng et al., 2007; Ronk, Korman, Hooke, & Page, 2013). The DASS has 21 items and participants respond to each item on a 4-point Likert-scale from 0 (did not apply to me at all) to 3 (applied to me very much or most of the time). The three subscales (i.e., depression, anxiety and stress scale), can be combined to a total scale score covering general distress used in the present study (Henry & Crawford, 2005; Osman et al., 2012). Within the current sample, the DASS general distress total scale score reached very good internal consistency, CR α = 0.937 (subscales: stress CR α = 0.890, depression CR α = 0.935, anxiety CR α = 0.837). Within the current sample, the DASS general distress total scale score was used to assess treatment outcome. Since symptom co-morbidity is high between anxiety disorders and depression, and thus treatments of anxiety disorders also reduce symptoms of depression (Cuijpers, Cristea, Weitz, Gentili, & Berking, 2016) and vice versa (Weitz, Kleboer, van Straten, & Cuijpers, 2018), we used the general distress scale of the DASS to fully cover treatment outcome (for a similar approach see von Brachel et al., 2019).

2.6. Procedure

Patients were contacted by the attending therapist prior to the beginning of their treatment. Once a patient signaled willingness to participate in the current study, he/she was contacted by a trained research assistant. Appointments with the patients were made between two probationary sessions prior to the beginning of the actual treatment and after the diagnostic session had taken place.

Prior to the beginning of the experimental session, participants gave...
informed consent to procedures. Participants were then seated in a dimly lit room and electrodes were attached. Then the conditioning task began. As part of a larger project, participants then performed a number of additional tasks (reported elsewhere). Before participants were dismissed, they filled in the DASS. After the experimental session, all participants received treatment-as-usual as routinely provided at our outpatient center. At the end of treatment, all patients were asked to fill in the DASS again.

2.7. Data analysis
2.7.1. Analysis of differences in fear acquisition and extinction between diagnostic groups

For subjective ratings, mixed model ANOVAs with the between subject factor diagnostic group (HC, DEP, ANX) and the within subject factors CS-type (CS-, CS+) and experimental phase (pre-acquisition, acquisition, extinction) were calculated. Significant main effects and interactions were followed-up with t-tests. The alpha level was set to 0.05. If necessary, Huyn-Feldt corrected degrees of freedom were calculated. For the skin conductance response, separate ANOVAs including the between subject factor diagnostic group (HC, DEP, ANX) and the within subject factors CS-type (CS-, CS+) for the pre-conditioning, fear acquisition and extinction phase were calculated. In addition, to assess learning curves, we added the factor time to each of the three ANOVAs. For the pre-conditioning phase, the factor time had three levels (i.e., the three trials for each CS in the pre-conditioning phase), for the acquisition and the extinction phase, the factor time had six levels (i.e., the six trials for each CS).

We additionally performed the same set of ANOVAs (i.e. for valence and US contingency ratings, as well as SCR data) including patients grouped according to their comorbidities. Within these analyses we did not find any additional differences between diagnostic groups. The results of this analysis can be found in the supplemental materials (see Supplemental Materials section 2 and Tables S1–S5).

2.7.2. Dimensional analysis of individual differences in fear acquisition and extinction

For subjective ratings, ANCOVAs with the within subject factors CS-type (CS-, CS+) and experimental phase (pre-acquisition, acquisition, extinction) were calculated. Instead of the between subject factor diagnostic group, the depression and anxiety subscales of the DASS were added as a covariate to the models. For the skin conductance response, separate ANCOVAs including the between subject factors CS-type (HC, DEP, ANX) and the within subject factors CS-type (CS-, CS+) for the pre-conditioning, fear acquisition and extinction phase were calculated. In addition, to assess learning curves, we added the factor time to each of the three ANCOVAs (see above). We again added depression and anxiety as a continuous covariate to all models. Significant main effects and interactions were followed-up with t-tests. The alpha level was set to 0.05. If necessary, Huyn-Feldt corrected degrees of freedom were calculated. For each significant effect, eta squared effect sizes ($\eta^2$) are reported.

2.7.3. Associations between fear conditioning/extinction and treatment outcome

Prediction of treatment outcome. Prediction of treatment outcome with markers of fear acquisition and extinction was performed using CS-difference scores (i.e., CS+ - CS-). For US-expectancy and self-reported valence the ratings to the CS+ given at the end of the respective experimental phase were subtracted from those to the CS- during the same phase. For the SCR, difference scores were calculated from the average of allSCRs to the CS- subtracted from the average of all SCR to the CS+ during the respective experimental phase (i.e. fear acquisition, fear extinction). To predict pre-post symptom severity treatment outcome, the post-treatment general distress score of the DASS was used as dependent variable. For DEP and ANX patients, two models were calculated respectively. For both models, pre-treatment symptom severity was entered as a predictor at Level 1. Then either CS-difference scores (i.e., SCR, US expectancy rating, valence rating) of acquisition (Model 1) or extinction (Model 2) were entered as predictors at level 2. To control for multiple collinearity within our regression models, variance inflation factors (VIF) were additionally calculated. Results indicate, that collinearity most likely did not influence the current results.

3. Results

3.1. Diagnostic-based analysis of individual difference in fear acquisition and extinction

3.1.1. Valence ratings

Of the initial participants who finished the conditioning/extinction task, data of 5 healthy control participant and 1 depressed patient were lost. The final sample for valence rating analysis thus comprised of $n = 245$ participants ($n = 53$ healthy controls, $n = 100$ depressed patients, and $n = 92$ anxiety patients).

ANOVA revealed a significant main effect for CS-type, $F(1,242) = 141.24, p < .001, \eta^2 = .369$, a significant main effect for experimental phase, $F(2,484) = 28.10, p < .001, \eta^2 = .104$, and a significant interaction CS x Experimental Phase $F(2,484) = 103.18, p < .001, \eta^2 = .299$ (See Table 1 and Fig. 1 for descriptive statistics).

Follow-up tests for this interaction indicate that there was no difference in valence ratings between the CS+ and the CS- after the pre-conditioning phase, $t(247) = 1.26, p = .211$. In contrary, the CS+ was perceived as more negative than the CS- after the acquisition, $t(245) = 14.09, p < .001$, and still after the extinction phase, $t(246) = 5.62, p < .001$.

Towards the CS+ alone, valence ratings were more negative after the acquisition than after the pre-conditioning phase, $t(246) = 12.95, p < .001$, and significantly more positive after the extinction phase than the acquisition phase, $t(246) = -10.66 p < .001$.

The CS- was perceived more positive after the acquisition than after the

| Table 1 |
| DASS scores pre and post treatment. |
| | Pre treatment | Post treatment |
| | M | SD | M | SD |
| DASS general distress | | | | |
| HC | 9.32 | 7.64 | - | - |
| DEP | 28.33 | 12.21 | 23.48 | 13.73 |
| DASS depression | | | | |
| HC | 23.67 | 10.96 | 17.70 | 10.90 |
| ANX | 2.32 | 2.49 | - | - |
| DASS anxiety | | | | |
| HC | 11.34 | 5.99 | 8.75 | 5.96 |
| ANX | 7.20 | 5.07 | 5.41 | 4.65 |
| DASS stress | | | | |
| HC | 1.66 | 1.89 | - | - |
| DEP | 5.62 | 4.30 | 4.81 | 4.43 |
| ANX | 6.46 | 4.47 | 4.05 | 3.40 |

2 Summary of fear acquisition and extinction effects: Participants displayed successful fear acquisition and fear extinction. During the pre-conditioning phase, SCRs, valence ratings, as well as US-expectancy ratings towards the CS- and the CS+ did not differ. After the subsequent acquisition phase, valence ratings were more negative, and US-Expectancy ratings were higher for the CS+ as compared to the CS-. Likewise, during the course of the acquisition phase SCRs towards the CS+ were significantly larger as compared to the CS-, indicating successful fear acquisition. During the course of the extinction phase, SCRs towards the CS+ decreased and finally did not differ from those towards the CS-. Likewise, Us-Expectancy and valence ratings towards the CS+ were significantly smaller after extinction as compared to ratings assessed after the acquisition phase.
In the preconditioning phase, \( t(246) = 4.81, p < .001, \eta^2_p = 0.397 \), a significant main effect for experimental phase, \( F(2,486) = 45.22, p < .001, \eta^2_p = 0.157 \), and a significant interaction for CS-type x Experimental Phase, \( F(2,486) = 99.43, p < .001, \eta^2_p = 0.290 \). Follow-up tests for this interaction indicate that there was no difference in US expectancy ratings towards the CS+ and the CS- after the preconditioning phase, \( t(247) = 0.07, p = .941 \). In contrary, US expectancy ratings were higher towards the CS+ than the CS- after the acquisition, \( t(247) = 14.52, p < .001 \), and still higher after the extinction phase, \( t(247) = 7.71, p < .001 \) (for descriptive statistics see Table 1 and Fig. 2).

Towards the CS+ alone, US expectancy ratings increased from the preconditioning to the end of the acquisition, \( t(247) = 14.63, p < .001 \), and significantly decreased from the acquisition to the end of the extinction phase, \( t(247) = -10.75, p < .001 \).

Towards the CS-, US expectancy ratings decreased from the...
conditioning phase to the end of the acquisition phase, \( t(247) = -2.35, p = .020 \), but did not differ between the acquisition and the extinction phase, \( t(247) = 1.00, p = .329 \).

There were no significant main effects of interaction including diagnostic group (i.e., main effect for diagnostic group, \( F(2,243) = 0.34, p = .712 \); CS-type x Diagnostic Group: \( F(2,243) = 1.57, p = .210 \); Experimental Phase x Diagnostic Group \( F(4,486) = 1.00, p = .409 \); CS-type x Experimental Phase x Diagnostic Group \( F(4,486) = 0.82, p = .511 \)).

### 3.1.3. Skin conductance response

Of the initial participants who finished the conditioning/extinction task, data of 1 healthy control participant, 19 depressed patients and 16 anxiety patients were lost due to equipment malfunction. The final sample for SCR analysis thus comprised of 82 healthy controls, 76 depressed patients, and 57 anxiety patients for CS-type, experimental phase and time.

**Preconditioning.** Results indicate that SCRs towards the CS+ and the CS- did not differ during the pre-conditioning phase, as indicated by a non-significant main effect for CS-type, \( F(1,212) = 1.86, p = .175 \). However, regardless of CS-type, SCRs decreased over the course of the pre-conditioning phase, as indicated by a significant main effect for time, \( F(2,424) = 69.51, p < .001 \), and a corresponding linear trend, \( F(1,212) = 91.41, p < .001 \). The interaction CS-type x Diagnostic Group, and Time x Diagnostic Group did not reach significance (See Fig. 3).

**Acquisition.** The ANOVA revealed a significant main effect for CS-type, indicating that responses towards the CS+ were larger than those towards the CS-, \( F(1,212) = 92.01, p < .001, \eta^2_p = .303 \). A significant main effect for time, \( F(5,1060) = 6.50, p < .001, \eta^2_p = .030 \), and a corresponding linear trend for time, \( F(1,212) = 17.59, p < .001, \eta^2_p = .077 \), indicate that responses generally declined during the course of the acquisition phase. However, in addition SCRs differed as a function of CS-type x Time, \( F(5,1060) = 8.66, p < .001, \eta^2_p = 0.039 \). Follow-up tests for this interaction revealed that SCR towards the CS+ and the CS- did not differ during the first trial of the acquisition phase, \( p = .428 \), while SCRs towards the CS+ were larger during all remaining acquisition trials (i.e., trials 2–6, all \( p < .001 \)). There were no other significant main effects or interactions.

**Extinction.** Analyses revealed a significant main effect for time, \( F(5,1055) = 9.43, p < .001, \eta^2_p = 0.043 \), and main effect for CS-type, \( F(1,211) = 27.84, p < .001, \eta^2_p = .117 \), as well as a significant interaction for CS-type x Time, \( F(85,1055) = 4.73, p < .001, \eta^2_p = 0.043 \). Follow-up t-tests of this interaction showed that responses towards the CS+ were larger than those towards the CS- during the first, \( p < .001 \), and the second trial of the extinction phase, but not during the third trial, \( p = .962 \). SCRs to the CS- were again larger than those towards the CS- during the fourth trial, \( p = .012 \), and finally did not differ between the CSs during the last two trials of the extinction phase (i.e. trial five, \( p = .121 \), and trial six, \( p = .984 \)).

### 3.2. Dimensional analysis of individual differences in fear acquisition and extinction

For the ANOVAs, the between factor group was replaced by dimensional measures of depression and anxiety respectively (i.e., DASS depression subscale, DASS anxiety subscale, correlation, \( r(250) = .594, p < .001 \)). The remaining factors included in the analyses did not differ and results including the factors CS-type, experimental phase and time were comparable to those obtained from the diagnosis-based analyses. Thus, for reasons of brevity, only significant interactions and main effects including depression or anxiety are reported.

For subjective valence ratings, a significant main effect emerged for anxiety, \( F(1,241) = 4.52, p = .034, \eta^2_p = 0.018 \), indicating that greater anxiety was associated with more negative evaluation of the CSs independent of CS-type (i.e., CS+ or CS-). We additionally found a significant main effect for depression, \( F(1,241) = 4.62, p = .033, \eta^2_p = 0.019 \), and a significant interaction for CS-type x Depression, \( F(1,241) = 8.91, p = .003, \eta^2_p = 0.036 \). Follow-up correlational analysis of this interaction indicate that higher depression scores were associated with more negative evaluation of the CS+, \( r(246) = .205, p = .001 \), but not the CS-, \( r(246) = .047, p = .467 \).

For US expectancy ratings, we found a significant interaction for CS-type x Anxiety, \( F(1,242) = 5.53, p = .019, \eta^2_p = 0.022 \). Follow-up correlational analysis indicate that higher scores on the anxiety subscale of the DASS were associated with enhanced US expectancy towards the CS-, \( r(246) = .146, p = .022 \), but not the CS+, \( r(246) = .006, p = .921 \). There were no additional significant interactions including anxiety or depression.

### 3.3. Prediction of treatment outcome

#### 3.3.1. Treatment effect

Data from \( n = 75 \) patients with depressive and \( n = 60 \) patients with anxiety disorders were available for analyses. The remaining patients either did not attend the post-treatment questionnaire assessment (\( n = 23 \)) or prematurely terminated their treatments (\( n = 27 \)). Data of 8 patients were lost due to technical reasons. Importantly, patients who dropped out did not differ from those who continued the study in their gender, \( \chi^2(1) = 0.48, p = .488 \), or age, \( F(1,193) = 0.10, p = .938 \), nor their depression, \( F(1,193) = 0.01, p = .946 \), anxiety, \( F(1,193) = 0.16, p = .694 \), or stress symptoms scores \( F(1,193) = 0.21, p = .648 \) (i.e., as assessed with the DASS).

During treatment, for all patients (depression and anxiety), general
distress declined from pre-to post-treatment, \( F(1, 133) = 35.36, p < .001, \eta^2_p = 0.210 \) (\( M_{\text{pre treatment}} = 26.26, SD_{\text{pre treatment}} = 11.86; M_{\text{post treatment}} = 20.91, SD_{\text{post treatment}} = 12.84, d = 0.51 \)). This decline was independent of diagnosis, as indicated by a non-significant interaction time x diagnosis, \( F(1, 133) = 0.37, p < .542 \). However, in general, patients suffering from depression had higher levels of general distress compared to patients suffering from an anxiety disorder, as indicated by a significant main effect for diagnostic group, \( F(1, 133) = 7.62, p = .007, \eta^2_p = 0.054 \) (see Table 1 for descriptive statistics).

### 3.3.2. Prediction of treatment outcome with markers of fear conditioning and extinction

**DepressedPatients.** Multiple stepwise regressions indicate that none of the markers of fear acquisition or extinction (i.e., self-reported valence, US contingency and SCR) significantly predict changes in post-treatment symptom severity in depressed patients. However, as expected, post-treatment symptom severity was predicted by pre-treatment symptom severity (see Tables 2 and 3).

**Anxiety patients.** As for depressed patients, post-treatment symptom severity was significantly predicted by pre-treatment symptom severity in anxiety patients. Moreover, stepwise multiple linear regression indicates that in anxiety patients, better US expectancy discrimination learning during the acquisition phase predicted lower post-treatment symptom severity (see Tables 4 and 5). In addition, we found that smaller differences in SCR between CS+ and CS- during the extinction phase, that is, better extinction learning predicted lower post-treatment symptom severity (Fig. 4 shows the scatter plots of the significant prediction effects found in the regression analyses).

### 4. Discussion

The current study compared fear conditioning and extinction learning between patients with anxiety and depressive disorders, as well as healthy controls. We calculated both differences between diagnostic groups, and also analyzed the data trans-diagnostically across the entire sample. In addition, the patients underwent disorder-specific CBT to test if markers of fear conditioning and extinction learning are specifically associated with the outcome of CBT for anxiety disorders.

All groups showed successful differential fear acquisition and extinction learning on both objective/physiological (i.e., the skin conductance response) and subjective (i.e., valence ratings, US expectancy ratings) response measures. We found only very limited evidence for differences between healthy controls and the two patient groups in both the trans-diagnostic, as well as the group-based analyses. In contrast, our results indicate that in anxiety only, better extinction learning predicted more favorable safety learning during the fear acquisition phase, and more favorable extinction learning during the subsequent extinction phase. In detail, higher SCRs during the fear acquisition phase, and more favorable extinction contrast, our results indicate that in anxiety only, better treatment effects found in the regression analyses.

Note: n.s. = non-significant, *p < .05; **p < .01; ***p < .001, Val = Valence, Diff = Difference Score CS + - CS-, Acq = Acquisition, CI = Confidence Interval, US Exp = US Expectancy, VIF = Variance Inflation Factor.

### Table 2
Prediction of residual post treatment general distress in DEP using markers of fear acquisition.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>CI</th>
<th>ΔR²</th>
<th>F₁</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DASS pre</td>
<td>0.78***</td>
<td>[0.57, 1.00]</td>
<td>.46</td>
<td>51.47(1.62)**</td>
<td></td>
</tr>
<tr>
<td>Val diff Acq</td>
<td>-0.05**</td>
<td>[-0.14, 0.04]</td>
<td>.04</td>
<td>14.32(4.82)**</td>
<td></td>
</tr>
<tr>
<td>US-Exp diff</td>
<td>-0.03[n.s.]</td>
<td>[-0.09, 0.04]</td>
<td>.04</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>SCR diff Acq</td>
<td>-19.06</td>
<td>[-48.28, 10.62]</td>
<td>.08</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: n.s. = non-significant, *p < .05; **p < .01; ***p < .001, Val = Valence, Diff = Difference Score CS + - CS-, Acq = Acquisition, CI = Confidence Interval, US Exp = US Expectancy, VIF = Variance Inflation Factor.

### Table 3
Prediction of residual post treatment general distress in DEP using markers of fear extinction.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>CI</th>
<th>ΔR²</th>
<th>F₁</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DASS pre</td>
<td>0.74***</td>
<td>[0.51, 0.97]</td>
<td>.41</td>
<td>41.24(1.60)**</td>
<td></td>
</tr>
<tr>
<td>Val diff Acq</td>
<td>-0.06^</td>
<td>[-0.08, 0.19]</td>
<td>.05</td>
<td>11.85(4.80)**</td>
<td></td>
</tr>
<tr>
<td>US-Exp diff</td>
<td>-0.05^</td>
<td>[0.04, 0.04]</td>
<td>.04</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>SCR diff Acq</td>
<td>29.13***</td>
<td>[-2.37, 60.62]</td>
<td>.04</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: n.s. = non-significant, *p < .05; **p < .01; ***p < .001, Val = Valence, Diff = Difference Score CS + - CS-, Acq = Acquisition, CI = Confidence Interval, US Exp = US Expectancy, VIF = Variance Inflation Factor.

### Table 4
Prediction of residual post treatment general distress in ANX using markers of fear acquisition.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>CI</th>
<th>ΔR²</th>
<th>F₁</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DASS pre</td>
<td>0.57***</td>
<td>[0.34, 0.79]</td>
<td>.35</td>
<td>25.31(4.80)**</td>
<td></td>
</tr>
<tr>
<td>Val diff Acq</td>
<td>0.07^</td>
<td>[-0.04, 0.02]</td>
<td>.02</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>US-Exp diff</td>
<td>-0.07*</td>
<td>[-0.13, -0.01]</td>
<td>.01</td>
<td>1.32</td>
<td></td>
</tr>
<tr>
<td>SCR diff Acq</td>
<td>-8.88^</td>
<td>[-34.07, 16.30]</td>
<td>.08</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: n.s. = non-significant, *p < .05; **p < .01; ***p < .001, Val = Valence, Diff = Difference Score CS + - CS-, Acq = Acquisition, CI = Confidence Interval, US Exp = US Expectancy, VIF = Variance Inflation Factor.

### Table 5
Prediction of residual post treatment general distress in ANX using markers of fear extinction.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>CI</th>
<th>ΔR²</th>
<th>F₁</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DASS pre</td>
<td>0.57***</td>
<td>[0.34, 0.79]</td>
<td>.35</td>
<td>25.31(4.80)**</td>
<td></td>
</tr>
<tr>
<td>Val diff Ext</td>
<td>0.02^</td>
<td>[-0.04, 0.08]</td>
<td>.01</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td>US-Exp diff</td>
<td>-0.04^</td>
<td>[-0.15, 0.08]</td>
<td>.08</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>SCR diff Ext</td>
<td>39.93*</td>
<td>[1.43, 78.44]</td>
<td>.08</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: n.s. = non-significant, *p < .05; **p < .01; ***p < .001, Val = Valence, Diff = Difference Score CS + - CS-, Acq = Acquisition, CI = Confidence Interval, US Exp = US Expectancy, VIF = Variance Inflation Factor.

towards the CS+ in a direct comparison to the CS- during the extinction phase, that is worse extinction learning, was associated with worse treatment outcome - that is higher post treatment symptomatology. In turn, less discrimination between the CS+ and the CS-
in terms of US-expectancy after the fear acquisition phase, was associated with less residual symptomatology after treatment. In other words, those patients who learned well that the CS- signaled safety whereas the CS+ displayed danger had a better treatment outcome.

As expected, no significant associations were found between any marker of fear acquisition or extinction learning and treatment outcome in depressed patients. To the best of our knowledge, this is the first study providing direct evidence for disorder-specific predictive validity of an extinction learning protocol for the outcome of cognitive behavior therapy in anxiety.

4.1. Prediction of treatment outcome

The current findings are in line with previous studies supporting the predictive validity of extinction learning for the outcome of exposure-based treatments of anxiety (review in Scheveneels et al., 2021). Indeed superior extinction learning as assessed with physiological responses has been shown previously to predict better treatment outcome in children and adolescents with various anxiety disorders treated with CBT (Waters & Pine, 2016) as well as non-clinically anxious adults receiving a one session exposure treatment (Forcadell et al., 2017). Thus, in sum our results extend these data and show predictive validity also for an adult patient sample being treated in a typical routine care setting with individual multi-session treatments. In addition, our study provides first evidence for predictive validity of extinction learning in a naturalistic, non-selected sample of patients with various anxiety disorders attending for treatment in an outpatient center. That is, none of our patients were rejected due to comorbidities and all patients received treatments-as-usual, as commonly applied in routine care at our center. The fact the current predictive associations emerged in a treatment-as-usual setting, where treatments are (1) typically less structured (von Brachel et al., 2019) and (2) often include less exposure (Pittig, Kotter, & Hoyer, 2019) than treatments in randomized controlled trials, support a rather robust association between extinction learning and treatment outcome. Moreover, as we did not find an association between extinction learning and treatment outcome in depression, our data confirms that this association is specific for treatment within the anxiety disorder spectrum. Both disorders share relevant symptomatology and thus CBT approaches for both disorders share relevant features (i.e., cognitive interventions, relaxation, contingency management, etc.), but differ in terms of exposure-based techniques, which are typically almost exclusively applied within anxiety treatments. Our data therefore supports and extend previous work showing that activation in emotion-related brain areas during extinction is associated with exposure-based treatments but not with a relaxation-based intervention (Lange et al., 2020). In sum, the current data thus provides further support for recent theoretical models on extinction learning as an experimental model for exposure-based treatments (overview in Craske, Hermans, & Vervliet, 2018) for anxiety disorders also in primary care settings under real-life treatment-as-usual conditions.

In addition to extinction learning, we found favorable US-expectancy
ratings (greater discrimination between the CS+ and the CS-) during fear acquisition to be predictive of better subsequent treatment outcome. Again, these findings were specific for patients with anxiety disorders. In detail, our results indicate that in anxiety patients only, more effective safety learning in terms of US-expectancy ratings (i.e., better discrimination of the CS+ and the CS-) during fear acquisition is predictive of better treatment outcome. Previous work suggest that anxiety might be associated with less favorable safety learning, (i.e., elevated fear responses towards the conditioned safety cue, CS-; (Duits et al., 2015; Lissek et al., 2008). Indeed, safety learning during a differential fear conditioning paradigm requires the acquisition of an inhibitory association between the CS- and the omission of the feared US (see for example Lonsdorf & Richter, 2017). Difficulties in building this inhibitory association most likely rely on the acquisition of false expectations towards the conditioned safety cue. Problematic expectancies towards ambiguous stimuli or situations have repeatedly been suggested to represent key features of anxiety disorders (e.g. Rief et al., 2015) and they probably contribute to elevated fear generalization (Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015), a process based on either perceptual or conceptual similarities between the CS+ and the CS- (for example: both CSs were Rorschach pictures in the current experiment). Excessive fear generalization has been suggested to one core mechanism underlying anxiety disorders (Dymond et al., 2015; Pittig, Treanor, LeBeau, & Craske, 2018). In this sense, elevated fear generalization/reduced safety learning contributes to the worsening of anxiety symptomatology and intensified avoidance behaviors (Dymond et al., 2015; Lissek, 2012; Wong & Pittig, 2020) through the spreading of acquired fear to other objects resembling the originally feared stimulus. In showing that more favorable safety learning/less pronounced generalization predicts more favorable treatment outcome, our data support previous reports on the beneficial effects of generalization training for the outcome of exposure-based interventions (Rowe & Craske, 1998) and theoretical models proposing inhibitory learning to be a key mechanism underlying fear exposure treatments (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). Furthermore, they suggest that safety learning or the ability to effectively discriminate between the threat-inducing properties of (perceptually or conceptually) similar stimuli may be a further disorder-specific mechanism moderating the outcome of exposure-based treatments within the anxiety disorder spectrum.

Finally, because they represent a convenient way to assess the different processes at work during fear acquisition or extinction (see Lonsdorf & Richter, 2017), we carried out our analysis on the predictive validity using CS+ - CS- - discrimination scores. To figure out, if the CS+ or the CS- has greater relative predictive power for treatment outcome, we additionally calculated regression models incorporating CS+ and CS- as separate predictors (see supplemental materials Tables S6-59). Interestingly, these analyses show that both, the CS+ and CS- are putatively involved in the significant prediction effects. From a methodological perspective, it has been previously suggested that differential conditioning paradigms (i.e. those incorporating a reinforced CS+ and non-reinforced CS-) are especially useful to disentangle interindividual differences in clinical contexts (c.f. Duits et al., 2015). In sum, our data support this assumption and shows that also the analyses of the predictive validity benefits from incorporating responding to both the CS+ and CS-.

4.2. Individual differences in fear acquisition and extinction learning

In contrast to group-differences in predictive validity for treatment outcome, our data shows no evidence for baseline differences between (group-based analyses) or across diagnostic groups (dimensional analyses) in differential fear acquisition and extinction learning. These findings are at odds with recent meta-analyses confirming impaired extinction learning in patients with anxiety disorders (i.e., greater fear responses towards the CS+ during extinction) as compared to healthy controls (Duits et al., 2015; Lissek et al., 2005). In addition, our data do not confirm previous reports on differences in fear acquisition or extinction learning in depression (Nissen et al., 2010; Otto et al., 2014). Rather, our data are in line with a recent study who did not find differences in fear acquisition or extinction in a large sample of patients with various anxiety disorders in comparison to healthy controls (Pöhlichen et al., 2020). Moreover, the Pöhlichen et al. (2020) study points to rather low effect sizes for extinction deficits in anxiety patients in previous meta-analyses (Duits et al., 2015), especially for studies using a rather unspecified electrical shock as US. Electrical shock is widely used in the conditioning literature (Lonsdorf & Richter, 2017) and was also used as US in the current study. In contrast, Duits et al. (2015) report considerably larger effect sizes for studies using fear specific US, an effect probably relying on enhanced individual motivational significance (Lang, Davis, & Ollman, 2000) provided by the individually fear-specific US. In line with this, previous research into anxiety disorders has demonstrated that specific fear-relevant cues in contrast to generally negative control cues (1) trigger larger defensive physiological responses (Hamm, Cuthbert, Globisch, & Vaitl, 1997; Wannemiller, Adolph, Joehn, Blackwell, & Margraf, 2017) (2) are accompanied by specific deficits in emotion regulation (e.g., Hermann et al., 2009; Paul, Simon, Endrass, & Kathmann, 2016), and are able to (3) trigger disorder specific interpretation biases (e.g., Amin, Foa, & Coles, 1998; In-Albon, Klein, Rinck, Becker, & Schneider, 2008; Woud, Zhang, Becker, McNally, & Margraf, 2014). Thus, taken together, specific fear-relevant stimuli seem to play a pivotal role in guiding anxiety related cognitive and emotional processes in the anxiety disorders spectrum. It might thus be worth investigating in future studies whether anxiety is indeed accompanied by a general deficit in extinction learning (e.g., Graham & Milad, 2011) or if extinction deficits might be largely pronounced or even restricted to individually significant, specific fear relevant stimuli.

In addition to the stimulus’ fear-relevance, future studies could consider to vary stimulus contingencies to enhance patient-control differences in conditioning experiments. Indeed, we used a 100% contingency making the CS+ - a rather unambiguous signal for the upcoming threatening consequence (i.e. the UCS). Previous studies found that unambiguous threat evoke rather indistinguishable responses among patients with anxiety disorders and healthy controls, while weakening the anxiogenic salience may more readily trigger patient-control differences (for a comprehensive discussion of this issue see, Lissek, Pine, & Grillon, 2006).

While we did not find individual differences in fear acquisition and extinction learning, the categorical, as well as the dimensional analytic approach indicate individual differences in the subjective evaluation of the conditioned safety cue, CS-. In detail, the group-based analysis indicates that valence ratings in response to the CS- presentations were more negative in patients (depression and anxiety). Supporting these data, the dimensional analysis revealed an association between more negative valence ratings of the CS- with higher scores on the depression and anxiety subscale of the DASS. In addition, the dimensional, but not the group-based categorical analysis showed that higher US-expectancy ratings towards the CS- were associated with higher scores on the anxiety, but not on the depression subscale of the DASS. Importantly, all effects were evident across experimental phases, and not restricted to the fear acquisition phase only. As already discussed above, individual differences in responding towards the CS- may represent individual differences in safety learning via inhibitory conditioning (CS- omission contingency) (Lonsdorf et al., 2017; Lonsdorf & Richter, 2017). During a typical fear conditioning study, these associations are established during the acquisition phase where the actual associative learning process takes place. On the contrary, during the extinction phase the retrieval of the previously learned association is paramount for the generation of appropriate responses towards the previously conditioned stimuli (Lonsdorf & Richter, 2017). Thus, since our ANOVA results (interaction for CS x DASS anxiety and depression scale) indicate that enhanced...
responding towards the CS- was not restricted to the acquisition phase only, it might be speculated that the sustained negative evaluation of the CS- not only represent unfavorable safety learning but also problematic retrieval of safety information for individuals experiencing elevated symptoms of anxiety and depression.

Interestingly, larger US-expectancy ratings towards the CS- were associated with higher scores on the anxiety scale of the DASS but not the depression subscale, again indicating the central a role of maladaptive expectancies within the anxiety disorders spectrum (Rief et al., 2015). The fact that we did not find differences in CS- expectancy ratings within our group-based analysis might indicate higher sensitivity of the trans-diagnostic (dimensional) approach, especially when investigating a naturalistic, non-selected sample yielding comorbidity rates between anxiety and depressive disorders. In general, effect sizes for individual differences were nonetheless rather small and thus the current effects should be interpreted with caution. However, they might serve as a starting point for future research.

4.3. Limitations

Our study comes with limitations. First, because we decided to recruit a naturalistic, ecologically valid sample of patients receiving treatment-as-usual in our outpatient center, we were unable to closely monitor the therapeutic process. That is, ecological validity of the sample and the treatment process most likely comes with limitations in treatment integrity. However, as already mentioned before, treatments in our center regularly follow published CBT manuals for the respective disorders (von Brachel et al., 2019).

Second, a considerable number of patients either terminated the treatment prematurely or did not attend the post-treatment questionnaire assessment. This renders the power of our analyses concerning prediction of treatment outcome considerably lower as compared to the analyses of baseline differences between diagnostic groups. However, the response rate is comparable to previous studies in a comparable setting (von Brachel et al., 2019). Moreover, because we recruited a large patient sample, the final N for the prediction of treatment outcome was still considerably high (n = 135), still allowing for sufficiently powered prediction analyses. This is also confirmed by a post-hoc power estimation demonstrating that our sample was sufficiently powered to detect medium to large effect sizes in multiple regression analyses (i.e., f² = 0.25, 1-β = 0.90, α = 0.05, number of predictors = 4). Nonetheless, it cannot entirely be ruled out that we failed to detect small effects in our data. However, importantly, patients who finished the study, and those who dropped out did not differ from each other in terms of gender, age, or their depression, anxiety or stress level. Thus, most likely the current effects were not affected by the current drop-outs.

Third, in contrast to previous studies (i.e., Ball, Knapp, Paulus, & Stein, 2017; Forcadell et al., 2017), we did not find an association of subjective ratings during extinction learning with subsequent treatment outcome. In general, emotional reactions can be measured on different response systems (i.e., subjective, physiological or behavioral responses) which do not necessarily show coherent response patterns. Indeed, especially between subjective and physiological readout measures, including for example the skin conductance responses and affective ratings, response coherence is often less pronounced as compared to behavioral and subjective emotional responses (for a comprehensive discussion see Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). In addition, some authors propose different learning processes underlying expectancy and affective learning respectively within conditioning studies (e.g., Hamm & Vaitl, 1996), with the former process relying more strongly on UCS-CS contingency awareness than the latter (Sevian, Beckers, & Kindt, 2012). Therefore, current recommendations include multiple response measurements in fear conditioning studies (Lonsdorf et al., 2017). Thus, in sum -, the different result patterns may have emerged due to different learning processes, leading to limited response coherence.

4.4. Summary and outlook

The current study demonstrates a striking dissociation between the predictive validity and the diagnosis-based individual differences of extinction and safety learning within the anxiety disorders spectrum. Confirming high specificity of the predictive validity of both processes for exposure-based CBT in the anxiety disorders spectrum, more favorable performance in both processes was associated with more favorable treatment outcome for patients with anxiety disorders, but not for patients with depression. These data underline the importance of fear-inhibitory learning processes as mechanisms of successful exposure-based treatments as well as the utility of experimental models for the investigation of treatment effects in exposure therapy for anxiety disorders.

In contrast, our data did not confirm pronounced differences between patients with depression or anxiety and healthy controls. If at all, our data supports small individual differences in safety learning which were statistically more pronounced in the dimensional analytic approach. This might indicate that a trans-diagnostic approach (Cuthbert, 2014) might be more sensitive to detect subtle effects in heterogeneous samples with high comorbidity rates within the internalizing disorders spectrum. However, effects were small and clearly further research is needed to disentangle the role of extinction and fear learning within the etiology of anxiety disorders. Within this endeavor, data driven approaches aimed at detecting distinct patterns of fear responding across individuals might be promising to take into account within group variance (Duits et al., 2021; Lonsdorf & Richter, 2017). In addition, future research may also profit from recent attempts to provide more psychometrically robust, transdiagnostic and dimensional clinical measures (Kotov et al., 2017) which may allow for a more ecologically valid assessment of psychopathology and overcome known issues with current nosologies (Michelini, Palumbo, DeYoung, Lutman, & Kotov, 2021).

In sum, the current study provides ample evidence for specific predictive validity of extinction and safety learning for the outcome of exposure-based cognitive behavior therapy in the anxiety disorders spectrum under real-life treatment-as-usual conditions. Thus, it confirms and extends current models rendering fear extinction and safety learning central mechanisms predicting the reduction of pathological fear during exposure-based treatments and useful experimental models for the investigation of treatment effects in exposure therapy for anxiety disorders.

CRediT authorship contribution statement

Dirk Adolph: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. Tobias Teismann: Resources, Writing – review & editing. Andre Wannemüller: Writing – review & editing. We would like to thank Nina Trimborn and Alina Schlacht for their help with performing the experiments. Jürgen Margraf: Conceptualization, Resources, Methodology, Supervision, Writing – review & editing.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2022.104229.


