Sometimes I feel the fear of uncertainty: How intolerance of uncertainty and trait anxiety impact fear acquisition, extinction and the return of fear

Adrian Wroblewski, Maike Hollandt, Yunbo Yang, Isabelle C. Ridderbusch, Anne Pietzner, Christoph Szeksa, Martin Lotze, Hans-Ulrich Wittchen, Ingmar Heing, Christof Koleske, Martin Lotzec, Hans-Ulrich Wittchen, e, Ingmar Heing, e, Dirk Adolph, Jürgen Margraf, Ulrike Lueken, Paul Pauli, Martin J. Herrmann, Markus H. Winkler, Andreas Orshol, Udo Dannlowski, Tilo Kircher, Jan Richter, Jan Richter, Jan Richter

a Department of Psychiatry and Psychotherapy, Center for Mind, Brain and Behavior - CMBB, Philipps-University Marburg, Germany
b Department of Psychology, University of Greifswald, Germany
c Functional Imaging Unit, Diagnostic Radiology and Neuroradiology of the University Medicine Greifswald, Germany
d Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany
e Department of Psychiatry and Psychotherapy, University Hospital, Ludwig Maximilians University Munich, Germany
f Translational Psychotherapy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
g Institute for Translational Psychiatry, University of Münster, Germany
h LVR-Hospital Essen, Department of Psychiatry and Psychotherapy, University of Duisburg-Essen, Germany
i Institute of Psychology, TU Darmstadt, Germany
j Mental Health Research and Treatment Center, Ruhr-University Bochum, Germany
k Institute of Clinical Psychology, Psychosomatics, and Psychotherapy, University Hospital Würzburg, Germany
l Department of Psychology, Humboldt-Universität zu Berlin, Germany
m Department of Psychology I, Biological Psychology, Clinical Psychology, and Psychotherapy, University of Würzburg, Germany
n Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité - Universitätmedizin Berlin, Germany corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berliner Institut für Gesundheitsforschung, Germany

ARTICLE INFO

Keywords:
- Fear conditioning
- Intolerance of uncertainty
- Trait anxiety
- Psychophysiology
- Dispositional negativity

ABSTRACT

It is hypothesized that the ability to discriminate between threat and safety is impaired in individuals with high dispositional negativity, resulting in maladaptive behavior. A large body of research investigated differential learning during fear conditioning and extinction protocols depending on individual differences in intolerance of uncertainty (IU) and trait anxiety (TA), two closely-related dimensions of dispositional negativity, with heterogeneous results. These might be due to varying degrees of induced threat/safety uncertainty. Here, we compared two groups with high vs. low IU/TA during periods of low (instructed fear acquisition) and high levels of uncertainty (delayed non-instructed extinction training and reinstatement). Dependent variables comprised subjective (US expectancy, valence, arousal), psychophysiological (skin conductance response, SCR, and startle blink), and neural (fMRI BOLD) measures of threat responding. During fear acquisition, we found strong threat/safety discrimination for both groups. During early extinction (high uncertainty), the low IU/TA group showed an increased physiological response to the safety signal, resulting in a lack of CS discrimination. In contrast, the high IU/TA group showed strong initial threat/safety discrimination in physiology, lacking discriminative learning on startle, and reduced neural activation in regions linked to threat/safety processing throughout extinction training indicating sustained but non-adaptive and rigid responding. Similar neural patterns were found after the reinstatement test. Taken together, we provide evidence that high dispositional negativity, as

https://doi.org/10.1016/j.ijpsycho.2022.09.001
Received 2 March 2022; Received in revised form 31 August 2022; Accepted 6 September 2022
Available online 15 September 2022
0167-8760/© 2022 Elsevier B.V. All rights reserved.
1. Introduction

Differentiating between threat and safety, as well as updating related memories in given situations are fundamental abilities for wellbeing. However, environmental cues are often uncertain and volatile with regards to threat and safety and therefore provoke heterogeneous interpretations between individuals. In the laboratory, the individual ability to adapt to changing demands of threat and safety under uncertainty can be modeled by fear conditioning and extinction paradigms. In human experimental studies, differential conditioning paradigms are frequently used which equally consider both facets (Lonsdorf et al., 2017).

During fear acquisition, one neutral stimulus (conditioned stimulus, CS+) but not another one (CS-) is paired with an aversive event (unconditioned stimulus, US) during fear acquisition training. This pairing provokes threat (to the CS+) and safety (to the CS-) learning and results in a defensive conditioned response (CR) selectively to the CS+ but not to the CS-. During extinction training, the CS+ (and also the CS-) is presented again without being followed by the US, now provoking safety learning and resulting in decreasing CRs. Extinction learning is hypothesized to stimulate new CS+/no-US associations without affecting the previously formed excitatory CS+/US associations, but being able to inhibit the retrieval of the latter (Milad and Quirk, 2012). These, however, remain prone to return under certain circumstances, e.g., after reexposure to the US (reinstatement; Haaker et al., 2014). Importantly, fear conditioning protocols can provoke varying degrees of uncertainty about newly learned threat/safety associations during both fear acquisition training and extinction training based on specific methodological properties. These include, e.g., the degree of similarity between conditioned threat and safety cues, the degree of (temporal) coincidence between conditioned cues and unconditioned stimuli, the rates of US reinforcement, the degree of instructed contingencies, or intermediate vs. delayed extinction (Lonsdorf et al., 2017; Lonsdorf and Richter, 2017).

Differences in adaptive behaviors and learning of threat and safety during fear conditioning are strongly associated with inter-individual differences in biological factors and personality traits. One of the most influential traits is the higher-order construct of “dispositional negativity” or “negative emotionality” (Barlow et al., 2014; Caspi et al., 2005; Markon et al., 2005; Shackman et al., 2016; Van Den Berg et al., 2014; Van den Bergh et al., 2021; Widiger and Oltmanns, 2017), which was hypothesized to result from increased reactivity to uncertain stressors (Shackman et al., 2016).

Dispositional negativity is usually broadly construed to include several transdiagnostic dimensions related to the experience of anxious or depressive mood. The dispositional propensity to interpret ambiguous situations as uncertain and to associate these uncertainties with a negative belief is summarized under the concept of Intolerance of Uncertainty (IU; Freeston et al., 1994). Individuals high in IU show biased uncertainty about safety, social evaluation, or health status, providing a risk for manifold maladaptive behaviors (Birrell et al., 2011; Carleton et al., 2012; Grenier et al., 2005; Lonsdorf and Merz, 2017; Morriss et al., 2016a, 2016b; Morriss et al., 2021a; Morriss et al., 2021b). Accordingly, IU was suggested to be a transdiagnostic risk factor for the development and maintenance of anxiety, stress-related and mood disorders (Carleton et al., 2012; Gentes and Ruscio, 2011; McEvoy et al., 2019; McEvoy and Mahoney, 2012), and is also increasingly discussed as a potential treatment target of mental disorders (Boswell et al., 2013; Einstein, 2014; McEvoy and Erceg-Hurn, 2016; Morriss et al., 2020; Oglesby et al., 2017; van der Heiden et al., 2012).

IU was shown to be highly correlated with (Furtado et al., 2021; Huggins et al., 2021; Sexton and Dugas, 2009) and reliably predicted by trait anxiety (TA; Jensen et al., 2016). TA is considered a stable personality predisposition to demonstrate elevated anxiety-related feelings, thoughts and behaviors (Milad et al., 2005; Most et al., 2006; Omura et al., 2005; Pujol et al., 2002), and is also established as a transdiagnostic risk factor for emotional disorders, such as the anxiety disorders spectrum (Chambers et al., 2004; Nordahl et al., 2019; Schmidt et al., 2008). Both, IU and TA, are related to dispositional negativity, but capture slightly different aspects. While TA is argued to strongly overlap with the tendency to experience negative affect (Barlow et al., 2014; Clark and Beck, 2011), IU rather captures the need for controllability and predictability (Carleton, 2016a).

In a recent study (Sjoerverman et al., 2020), a latent factor “negative emotionality” (based on questionnaire scores for IU, TA, and neuroticism) predicted reduced fear conditioning, as reflected by decreased CS+/CS- discriminations in skin conductance responses (SCR), auditory startle blinks and subjective ratings. Reduced CS discrimination resulted from elevated responses to the CS-, which suggests deficient safety learning processes (Laing and Harrison, 2021). These results may support the notion that a higher-order construct might be indeed informative and able to explain superordinate associations with fear learning. An integrative perspective of personality traits might also contribute to a better understanding of the hitherto heterogeneous findings regarding the relationship of specific personality facets, such as IU and TA, and deviations in fear conditioning (see Lonsdorf and Merz, 2017 and Morriss et al., 2021b for respective reviews). Moreover, on a neural level, previous studies found dispositional negativity to be associated with altered activation patterns within regions associated with fear processing and regulation, such as the amygdala, hippocampus, anterior insula (aINS), bed nucleus of the stria terminalis (BNST), dorsal anterior cingulate cortex (dACC), orbitofrontal cortex and periaqueductual grey (Avery et al., 2016; Brinkmann et al., 2018; Calder et al., 2011; Cavanna and Shackman, 2015; Fox et al., 2015a; Fox et al., 2015b; Fox and Kalin, 2014; Hur et al., 2019; Shackman et al., 2011; Shackman and Fox, 2016).

More specifically, IU has been identified to modulate fear conditioning mechanisms, but it is still unclear to what extent. With regards to initial learning during fear acquisition, results have been mixed so far (Morriss et al., 2021b). During extinction training, IU seems to be associated with impaired extinction learning as indicated by increased SCR responses and neural activation in the amygdala (Morriss, 2019;
Morris et al., 2015; Morris et al., 2016a; Morris et al., 2019a; Morris et al., 2021a; Morris and van Reekum, 2019; Wake et al., 2020). However, effects of IU were mainly evident during late trials of extinction training or across the whole extinction training phase, but not during early trials.

TA is one of the most frequently addressed factors in research on individual differences in human fear conditioning (Lonsdorf and Merz, 2017; Morris et al., 2021b). However, it is still unclear whether and how TA affects underlying mechanisms of differential fear conditioning (Lonsdorf and Merz, 2017). For example, for fear acquisition training, early studies found a positive association of high levels of TA with psychophysiological markers, i.e., SCR (Indovina et al., 2011) and fear-potentiated startle (FPS; Gazendam et al., 2013), as well as neural activation in the amygdala (Indovina et al., 2011). Sjouwerman et al. (2020) found the opposite pattern within psychophysiological markers, but demonstrated supporting evidence for a positive link between TA and differential amygdala activation during fear acquisition training. Others did not find any associations between TA and either outcome measure of fear learning (Arnaudova et al., 2013; Chin et al., 2016; Joos et al., 2012; Klingelhöfer-Jens et al., 2021; Martínez et al., 2012; Morris et al., 2016a; Sehmeyer et al., 2013; Torrents-Rodas et al., 2013). As for extinction training, previous studies suggested associations between TA and differential conditioning in FPS (Gazendam et al., 2013), and amygdala responding (Barrett and Armony, 2009; Sehmeyer et al., 2011). However, in their meta-analysis including results of 18 different experiments, Morris et al., 2021a did not find an association between TA and SCR. Yet, these results were again dependent of the phase of extinction training (early vs. late) and experimental design (immediate vs. delayed extinction). Similar inconsistencies have been reported for a link between trait anxiety and reinstatement effects (Gazendam et al., 2013; Kindt et al., 2009; Kindt and Soeter, 2013; Martínez et al., 2012; Soeter and Kindt, 2010).

Taken together, previous findings do not yet integrate into a bigger picture about the effects of dispositional negativity, i.e., IU and TA as the main constructs of interest in fear conditioning research, on fear and safety learning mechanisms. This heterogeneity may derive from several, mainly methodological, limitations in earlier studies (see Lonsdorf and Merz, 2017 for review), for example 1) underpowered sample sizes, potentially overlooking small effects, 2) differences in experimental design, such as the reinforcement rate during fear acquisition training altering experienced uncertainties, or differences in extinction training procedures (immediate vs. delayed extinction training), 3) focusing on single outcome measures and thereby not being able to draw conclusions about general effects across them, or 4) considering only a single construct of negative affect instead of using compounds of strongly correlated constructs such as IU and TA. Finally, possible effects of individual differences on fear learning might be partially confounded (Lonsdorf and Richter, 2017).

Against this background, the purpose of the present study was to provide a systematic and integral investigation of time-sensitive fear conditioning mechanisms concerning dispositional negativity, assessed with self-reported IU and TA, reflecting two closely related dimensions. Therefore, we analyzed and expanded a large multimodal dataset previously collected in Ridderbusch et al., 2021, using a fear conditioning protocol optimized to investigate the specific effects on fear extinction learning with increasing clinical translation ability (see Hollandt et al., 2020 for a detailed discussion). The paradigm comprises instructed fear acquisition training (contingency between CS+ and US but no reinforcement rate were instructed) to ensure a well-established conditioned fear response prior to extinction training with low variance between participants. Twenty-four hours later, delayed extinction training and a subsequent return of fear test (i.e., reinstatement) followed. Importantly, before extinction no explicit instructions about US presentations or possible associations between presented CSs were given, resulting in an ambiguous state of US uncertainty. This was evidenced by lacking discriminations between CS+ and CS- on autonomic and defensive reflex measures during initial fear extinction, that was clearly observed during late trials of preceding acquisition training in this paradigm (Hollandt et al., 2020). This effect was based on increased physiological responding to both CS+ and CS- and went along with an increase of US expectancy ratings for the CS- but not the CS+. Thus, the newly developed paradigm is excellently suited to test for the general effects of dispositional negativity in threat and safety learning under experimental conditions of high uncertainty. To achieve a comprehensive picture of possible involved processes we assessed CRs by frequently used multimodal outcome measures in fear conditioning research (subjective ratings, SCRs, FPS, and BOLD fMRI). Due to the exploratory nature of the present analyses and the heterogeneity in the existing literature, we did not derive specific hypotheses about the direction of the effects, but expected differences between two clusters of participants with low vs. high IU/TA scores, especially during high levels of uncertainty during initial extinction training.

2. Materials and methods

Data used for the current study were acquired between 04/2016 and 04/2019 as part of the national research consortium “Providing Tools for Effective Care and Treatment of Anxiety Disorders” (PROTECT-AD). The overall aims of PROTECT-AD were to investigate the clinical role and underlying neural, psychophysiological, and (epi-)genetic mechanisms of exposure-based CBT (Heimig et al., 2017; Pittig et al., 2021). For the present paper, we analyzed data from healthy control participants that were collected in addition to patients studied during a fear extinction paradigm (Hollandt et al., 2020; Ridderbusch et al., 2021).

2.1. Participants

Healthy control participants from six sites in Germany with no 12-month histories of medical or mental illnesses as verified by the computer-assisted clinical version of the Composite International Diagnostic Interview (CIDE; Essau and Wittchen, 1993; Reed et al., 1998; Robins et al., 1988; Wittchen, 1994) were invited to take part in a two-day laboratory/MRI study. Due to analyses of (epi-)genetic effects on exposure-based CBT in PROTECT-AD, only participants with European descent were included in the study. The total sample included n = 155 participants with available questionnaire data. All those participants were investigated on day 1 (fear acquisition training) in the physiological lab and completed the study on day 2 (extinction training and reinstatement test) either in the physiological lab again or in the MRI scanner. A total of 15 participants did not attend day 2 measurements resulting in n = 140 participants with quality-controlled data sets investigated either in the physiological lab (n = 47) or in the MRI environment (n = 93). The quality control process for the MRI data was already described elsewhere (Ridderbusch et al., 2021). All participants had normal or corrected to normal visual acuity and were totally naive to the experiment beforehand. Based on the questionnaire data participants were divided into groups (low vs. high IU/TA) as detailed below. Table 1 summarizes demographic characteristics of the overall sample and the two sub-samples as a function of IU/TA groups, respectively.

2.2. Intolerance of uncertainty and trait anxiety scales

All invited participants were asked to complete a set of questionnaires prior to the experiment on study day 1 including the State-Trait Anxiety Inventory (STAI; Spielberger, 1983) and the German translation of the Intolerance of Uncertainty Scale (IUS; Buhr and Dugas, 2002; Freeston et al., 1994; Gerlach et al., 2008). The trait scale of the STAI (STAI-T) includes 20 self-reported items with statements, such as “I...
Taking into account the overlapping distribution of the IUS and STAI-T scales, we performed a multivariate K-means cluster analysis with a maximum of 25 iterations to force 2 separate groups (Fig. 1 for a scatterplot). Given the limited overall sample size, we decided on 2 groups to ensure sufficiently large groups for the planned analyses in the sub-groups, as well as using the sum scores of IUS and STAI-T instead of single item values. In general, cluster analyses can differentiate between groups based on similarity in various characteristics and patterns (Hennig et al., 2016; Klimberg et al., 2017). The goal is to minimize intra-cluster differences between samples while at the same time maximizing inter-cluster differences. Therefore, the comparison between the identified groups explicitly enables an optimal analysis between people who empirically differ maximally on the latent factor (estimated from the two questionnaires). In future applications, these clusters can be used, for example, to fit different learning models and identify further specific characteristics within the groups. Table 1 shows the respective number of participants.

### 2.4. Materials

All stimuli were presented using the Presentation software (version: 17.2, Neurobehavioral Systems, Albany, CA, http://www.neurobs.com/).

#### 2.4.1. Electrotactile stimulus

For the experimental procedure, we used an electrotactile stimulus as an unconditioned stimulus (US), which was administered to the forearm of the non-dominant hand using MRI-compatible reusable cup electrodes (10 mm silver, Medical Products, Wiesbaden) and specially produced electrode gel. The US consisted of a train of 125 and 100 single 5 ms electric pulses (625 and 500 ms total stimulus duration) for the
assessment in the physiological lab and MRI environment, respectively, and was generated by a constant current stimulator (DS7A, Digitimer, Medical Products, Wiesbaden). Prior to fear acquisition training and after the pre-conditioning phase on day 1, US intensity was individually calibrated using a standardized shock work-up procedure (Heitland et al., 2013) aiming at an unpleasant, but still tolerable and not painful intensity. For day 2, the identical electrode position and US intensity were used.

2.4.2. Auditory stimulus
A 50 ms burst of white noise with an intensity of 95 dB(A) (rise/fall <1 ms) served as a startle-eliciting probe and was presented binaurally over Sennheiser ARG K66 headphones (Sennheiser, Wedemark, Germany) either 4.5 or 5 s after CS onset and 2, 3, 4, 5, or 6 s after inter-trial-interval (ITI) onset. Six startle probes were presented alone before pre-conditioning on day 1 and before re-acquisition on day 2 in case of a respective assessment in the physiological lab to allow for startle habituation and to ensure a robust baseline of blink magnitudes (Blumenthal et al., 2005).

2.4.3. Visual stimuli
Two background-colored pictures of male neutral faces (from the Psychological Image Collection at Stirling; http://pics.stir.ac.uk; Duits et al., 2017) served as conditioned stimuli (CS). CSs were presented for 6.2 s (physiological lab) or 6 s (MRI) followed by an ITI (white fixation cross presented on a black screen) of 6–10 s. One of the two pictures, referred to as CS+, repeatedly co-occurred with the US (5.5 s after CS+) onset during fear acquisition training on day 1 and re-acquisition trial on day 2, while the other one (i.e., the CS-) did not. The allocation of stimuli to CS+ and CS- was counterbalanced between participants.

2.5. Experimental procedure
The experimental procedure was already described in detail elsewhere (Hollandt et al., 2020; Ridderbusch et al., 2021), hence we will only give a short overview here. See Fig. S1 in the supplement for a detailed description. To test for both CS+ and CS- on a continuous numeric rating scale (0–100 %, –5 [negative valence] – 5 [positive valence], and 0 [low arousal] - 10 [high arousal], respectively) in the laboratory environment (day 1 and day 2 non-fMRI sample). In the MRI environment subjective ratings of US expectancy (0–100 %), valence (0 [negative valence] – 100 [positive valence]) and arousal (0 [low arousal] - 100 [high arousal]) were assessed on a non-continuous 10 %-stepped rating scale. In the laboratory, US expectancy ratings were performed on a trial-by-trial basis, while valence and arousal were rated block-wise (once at pre and post-acquisition training, pre and post re-acquisition, after 10 and 20 extinction trials, post RI and post RIT). In the MRI environment, ratings, including US expectancy, were performed block-wise at the same times. Note that US expectancy ratings were always related to the upcoming CS, while ratings of arousal and negative valence were always presented subsequently to the CS. Furthermore, arousal and negative valence were not rated prior to the re-acquisition trial in the MRI environment. Further details can be found in Hollandt et al., 2020 and Ridderbusch et al., 2021.

2.6. Subjective ratings
Subjective ratings were assessed to indicate US expectancy, subjective valence, and arousal for both CS+ and CS- on a continuous numeric rating scale (0–100 %, –5 [negative valence] – 5 [positive valence], and 0 [low arousal] - 10 [high arousal], respectively) in the laboratory environment during which both CSs were presented again 20 times each without US pairing in the reinstatement test phase. After one re-acquisition trial (one CS+ followed by the US), extinction training was performed during which both CSs were presented again 20 times each without any US pairing. Subsequently, a return of fear manipulation took place where the US was presented three times while the screen remained black (reinstatement; RI). After RI, again both CSs were presented 10 times each without US pairing in the reinstatement test phase (RIT) and final re-extinction phase.

2.7. Verbal instructions
Verbal instructions for the participants were recently shown to have an impact on learning behavior during fear conditioning and therefore need careful consideration (Mertens et al., 2018, 2021). Prior to the acquisition training on day 1, participants were explicitly instructed about CS+/US contingency, but not the reinforcement rate, to ensure a robust fear response. On the second day, participants were instructed that a shock “may occur again” during the experiment without any further information.

2.8. Data acquisition and response definition
2.8.1. Skin conductance response (SCR)
SCR data from the hypothenar muscle were measured via self-adhesive Ag/AgCl surface electrodes (8 mm diameter, E224A, Warwick, RI; filled with isotonic 0.5 M sodium chloride electrode gel) placed on the palmar side of the non-dominant hand. Data were recorded with a sampling rate of 10 Hz by a Coulbourn S71-22 skin conductance coupler (Allentown, PA) providing a constant voltage of 0.5 V. SCRs were scored as the first response within a 0.90–4.00 s time window, following stimulus onset for CS and US, respectively, using an in-house software (Globisch et al., 1993), which can be made available upon request. Trials in which no SCR could be detected were scored as zero responses. In contrast, missing values during trials during which no reaction could be quantified due to recording artifacts (e.g., electrode malfunctions) were replaced individually for each subject by the overall SCR of this subject over all trials of the respective stimulus during the experiment. Base 10 logarithms for each value were then computed to normalize the distribution. To reduce interindividual variability of the SCR not related to the conditioning and extinction tasks of the experiment, the log values were range-corrected (division of individual score by the participants’ maximum response within all CS and US trials (Lykken and Venables, 1971)).

2.8.2. Startle blink magnitudes
The eye-blink component of the startle reflex was recorded via facial electromyography (EMG) using two Ag/AgCl surface electrodes (4 mm diameter, F-E9-60, Warwick, RI), filled with electrolyte paste (GE Medical Systems Milwaukee, WI), and a Coulbourn S75–01 bioamplifier (Allentown, PA), as well as a 400 Hz Kemo-VFB-8-03 low pass filter (Kemo, Darford, UK). A digital sampling rate of 1000 Hz was applied 100 ms prior to acoustic startle probe onset and lasted for 100 ms after acoustic startle probe onset. Raw data were filtered offline using a 60 Hz
highpass filter and were rectified and integrated (time constant: 10 ms) by a digital filter. Data were semi-automatically scored offline by using an in-house algorithm (Globisch et al., 1993) that identified latency of blink onsets and peak amplitudes. The time window for startle response was defined between 20 and 120 ms after startle probe onset and the magnitude had to peak within 150 ms after onset. If eyeblinks were not detectable, respective trials were scored as zero responses. Trials with excessive baseline activity, recording artifacts (e.g., electrode malfunctions), and spontaneous eyeblinks outside the latency window were treated as missing and therefore rejected. All participants met the 80% criterion for valid responses and could be included in statistical analysis. The missing values were replaced individually for each subject by the overall mean blink response magnitude of this subject over all trials of the experiment. Each response of each participant was then standardized and converted to t-scores [50 ÷ (x × 10)] to control for possible confounding effects of high inter-individual differences in baseline amplitude.

2.8.3. fMRI acquisition parameters
MRI data were acquired on 3 T MRI systems (3 × Siemens TrioTim, 1 × Siemens Verio, 1 × Siemens Prisma, 1 × Siemens Skyra, Erlangen) with 12-channel head coils. High-resolution (1x1x1mm3) T1-weighted anatomical images were acquired using a 3D magnetization prepared rapid gradient echo sequence (TE = 2.26 ms, TR = 1900 ms, inversion time (TI) = 900 ms, flip angle 9°, matrix size 256 × 256 voxels, slice thickness 1.0 mm, FOV = 256 mm, 176 slices) with identical settings in all centers. Functional images were obtained using a T2*-weighted gradient-echo planar imaging sequence sensitive for the BOLD contrast (TE = 30 ms, TR = 2000 ms, flip angle 90°, matrix size 64 × 64 voxels, voxel size 3.6 × 3.6 × 4.0 mm, slice thickness 4.0 mm, inter-slice gap 0.4 mm, field of view (FOV) = 230 mm, 33 slices, ascending phase encoding direction, T > C acquisition orientation, sequential acquisition; due to limitations in technical compatibility, a TE = 29 ms had to be used at the Siemens Prisma and at the Siemens Verio only 31 slices were recorded). Slices were positioned trans-axially parallel to the intercommissural (AC-PC) plane and tilted 20° to reduce magnetic susceptibility artifacts in prefrontal areas and cover the whole brain. In total, 590 volumes were collected.

2.9. Data reduction and analysis

2.9.1. Subjective rating and physiological data
Due to missing data due to technical faults or problems with physiological data recording, subjective rating data, startle responses, and SCR were available on day 1 for only 143, 138, and 131 participants, respectively. On day 2, startle responses and SCR were available for 45, and 38 participants, respectively. We identified startle missings in 629 (8.77 %) and 68 (1.47 %) trials for day 1 and day 2, respectively. The day 1 data (pre-conditioning phase and fear acquisition training) were analyzed using repeated-measures ANOVAs with Stimulus (CS+: vs. CS-: and - in the case of startle - vs. ITI) and Block (in case of US expectancy ratings and physiological outcomes two trials per block) as within-subject factors and Group (low vs. high IU/TA group) as between-subject factor. To harmonize with the fMRI data analysis, blocks were averaged over 5 trials for US expectancy ratings and physiological outcomes for day 2 data during extinction training. To analyze the effect of reinstatement, Block included the respective last response before and after the US alone presentations.

2.9.2. fMRI data analysis

2.9.2.1. Preprocessing. fMRI scans were already preprocessed and described earlier (Ridderbusch et al., 2021), so we will provide only a short description here. Data were analyzed using standard routines of the Statistical Parametric Mapping 12 (SPM) software (www.filion.uc.ac.uk/spm), based on Matlab R2009b (version 7.9.0; MathWorks). The preprocessing pipeline comprised co-registration to the anatomical T1 image, spatial realignment, and normalization to the MNI template, and an iterative spatial smoothing (Friedman et al., 2006) with a target kernel of 8 mm isotropic Gaussian filter.

2.9.2.2. First-level analysis. An event-related design, modelling each CS type, the US, and subjective rating phases was employed using a single subject first-level General Linear Model. Other than in our previous analyses (Hollandt et al., 2020; Ridderbusch et al., 2021), we decided to provide more time-sensitive analyses by subdividing the experiment into blocks of 5 trials each (as compared to blocks of 10 trials earlier), resulting in 14 experimental regressors and 6 nuisance regressors to account for movement-related noise, convolved with the hemodynamic response function. Single parameter estimates of the baseline t-contrasts for the Cs (CS+:, CS-) in the different blocks were calculated on the first level and then passed on to the group analysis.

2.9.2.3. Second-level analysis. At the group level, we conducted a flexible factorial whole-brain analysis including the 12 baseline contrasts (CS+: vs. implicit baseline and CS-: vs. implicit baseline in 6 blocks, respectively), as well as covariates of no interest for center, age (mean-centered), gender, and smoking. The statistical threshold was set to p < .05, but to account for multiple comparisons, we used a Monte Carlo simulation (Slotnick, 2017) at threshold p < .005 with a minimum cluster size of 175 contiguous voxels, identically to our previous analyses (Ridderbusch et al., 2021). For the estimation of spatial autocorrelation needed for the Monte Carlo simulation, we used the freely- available img_xcorr script (https://www2.bc.edu/sd-slotnick/scripts.html).

Contrasts of interest included effects of Stimulus (t-tests for CS+ > CS- and CS+ < CS-) during the whole extinction training, as well as during all blocks of extinction training, RIT and re-extinction. Main effects of Group during extinction training, RIT and re-extinction were assessed by F-test with the factor Group, as well as the two-way interaction with the factors Stimulus X Group. Furthermore, we used time-sensitive contrasts to investigate changes in the course of the experiment (three-way interactions with the factors Stimulus X Group X Block). For visualization purposes, we extracted mean beta values for the 12 baseline contrasts from specific regions of interest using the MarsBar toolbox implemented in SPM (Brett et al., 2002). Betas were extracted from whole clusters significantly activated at the corrected statistical threshold. Significant clusters were labelled based on the Neuromorphometrics atlas (Neuromorphometrics, Inc.) and complemented with functional labels where appropriate.

3. Results

3.1. Preconditioning phase and fear acquisition training (day 1)
Fig. S2 in the supplements shows the day 1 means for all subjective ratings of US expectancy (Fig. S2 A), valence (Fig. S2 B) and arousal (Fig. S2C) to CS+ and CS- separately for blocks of trials averaged across two trials in the overall investigation sample. Fig. 2 shows mean responses for FPS and SCR during day 1. During the pre-conditioning phase CS+ and CS- did not differ according to US expectancy ratings, valence and arousal ratings, startle blink magnitudes, and SCR magnitudes, respectively, in both groups. As expected, robust fear acquisition effects were observed after instruction and during acquisition training as
indicated by higher responses for CS+ relative to CS- in US expectancy ratings \( F(1,141) = 1095.13, p < .001, \eta^2 = 0.89 \), valence ratings \( F(1,141) = 89.42, p < .001, \eta^2 = 0.39 \), arousal ratings \( F(1,141) = 186.61, p < .001, \eta^2 = 0.57 \), startle blink magnitudes \( F(1,136) = 151.29, p < .001, \eta^2 = 0.53 \), and SCR magnitudes \( F(1,129) = 124.72, p < .001, \eta^2 = 0.49 \). Again, overall CS discrimination did not significantly differ between groups. However, habituation of SCRs to the CS+ but not to the CS- was delayed in the high IU/TA group (Fig. 2 D) relative to the low group (Fig. 2 C) resulting in a significant quadratic trend for Group X Stimulus X Block \( F(1,129) = 4.87, p < .05, \eta^2 = 0.04 \).

### 3.2. Extinction training (day 2)

#### 3.2.1. Subjective ratings

Fig. S3 in the supplements shows the day 2 means for all subjective ratings of US expectancy (Fig. S3 A-B), valence (Fig. S3 C-D) and arousal (Fig. S3 E-F) to CS+ and CS- separately for the physiology lab \( n = 47 \) and MRI samples \( n = 93 \). Replicating previous results, US expectancy ratings substantially increased for the CS- but not the CS+ from final fear acquisition training to initial extinction training in the physiological sub-sample (Stimulus X Block F(1,45) = 27.59, \( p < .001, \eta^2 = 0.38 \). Also, valence (Stimulus X Block F(1,45) = 13.96, \( p < .001, \eta^2 = 0.24 \) and arousal (Stimulus X Block F(1,45) = 39.39, \( p < .001, \eta^2 = 0.47 \) ratings changed differentially indicating a stronger decrease (valence) or increase (arousal), respectively, of subjective responding during CS- relative to CS+. Overall, here we found no differences between IU/TA groups. However, after the re-acquisition trial and prior to extinction training and compared to the CS-, the CS+ was still rated in both sub-samples to be more negative (physiological sub-sample: F(1,45) = 13.61, \( p < .001, \eta^2 = 0.23 \); MRI sub-sample: F(1,91) = 16.13, \( p < .001, \eta^2 = 0.15 \) and arousing (F(1,45) = 7.35, \( p < .01; \eta^2 = 0.14 \); F(1,91) = 55.53, \( p < .001; \eta^2 = 0.38 \), and was associated with higher US expectancies (F(1,45) = 34.59, \( p < .001, \eta^2 = 0.44 \); F(1,91) = 188.62, \( p < .001; \eta^2 = 0.68 \)) with, again, no significant differences between IU/TA.

During following extinction training, in both sub-samples the valence ratings continuously increased (getting more positive) for both CS+ and CS- (linear trend Block: F(1,45) = 10.40, \( p < .01, \eta^2 = 0.19 \); F(1,91) = 4.09, \( p < .05, \eta^2 = 0.04 \), but increase was stronger for CS+ in the physiological sub-sample (linear trend Stimulus X Block: F(1,45) = 4.68, \( p < .05, \eta^2 = 0.09 \), but not in the MRI sub-sample. Also, arousal ratings continuously decreased in both sub-samples for both CS+ and CS- (linear trend Block: F(1,45) = 21.74, \( p < .001, \eta^2 = 0.33 \); F(1,91) = 51.73, \( p < .001, \eta^2 = 0.36 \) with, however, a delayed decrease for the CS- relative to the CS+ (quadratic trend Stimulus X Block: F(1,45) = 10.97, \( p < .01, \eta^2 = 0.20 \); F(1,91) = 3.88, \( p < .05, \eta^2 = 0.04 \). Overall, arousal ratings during extinction training were higher for both CS+ and CS- for the high IU/TA group as compared to the low group in the MRI sub-sample (Group F(1,91) = 4.33, \( p < .05, \eta^2 = 0.05 \), but not in the physiological sub-sample. Also, the trial-by-trial measured overall US contingency ratings in the physiological sub-sample continuously decreased over the four blocks (each averaged across five trials; linear trend Block F(1,45) = 187.58, \( p < .001, \eta^2 = 0.81 \) with stronger decreases for the CS+ as compared to the CS- (linear trend Stimulus X Block F(1,45) = 29.16, \( p < .001, \eta^2 = 0.39 \). The same was observed for the three block-wise US contingency ratings in the MRI sub-sample (linear trend Block F(1,91) = 70.16, \( p < .001, \eta^2 = 0.44 \); linear trend Stimulus X Block F(1,91) = 13.07, \( p < .001, \eta^2 = 0.13 \). Importantly, IU/TA groups did not significantly differ with regard to the subjective rating changes during the extinction training.

#### 3.2.2. Physiological responses

During the first block of extinction training (trials 01–05), we found stronger startle responses during CS+ relative to CS- for the high IU/TA group, but not for the low IU/TA group (Group X Stimulus F(1,43) = 4.36, \( p < .05, \eta^2 = 0.09 \); Fig. 3 A and B). In the high IU/TA group, startle responses decreased during extinction training comparable for both CS+ and CS- suggesting no differential learning and resulting in a
robust CS startle discrimination throughout the extinction training phase (Stimulus $F(1,17) = 25.72$, $p < .001$, $\eta^2_p = 0.60$). In contrast, startle started to strongly discriminate between CS+ and CS− during extinction block 2 (trials 06–10) in the low IU/TA group due to a stronger decrease of startle responses during CS− relative to CS+ (Fig. 3 B). Differential startle responses decreased again during the last two blocks (trials 11–20) resulting in a significant quadratic trend (Group X Stimulus X Block $F(1,43) = 26.34$, $p < .001$, $\eta^2_p = 0.38$) comparable between both IU/TA groups. However, although not statistically significant, overall fear-potentiated startle tended to be stronger in the high IU/TA group relative to the low group although not statistically significant, overall fear-potentiated startle (Stimulus $F(1,17) = 5.69$, $p < .05$, $\eta^2_p = 12$). Fear-potentiated startle to the CS+ (relative to ITI) continuously decreased during extinction training (linear trend Stimulus X Block $F(1,43) = 26.34$, $p < .001$, $\eta^2_p = 0.38$) comparable between both IU/TA groups. However, although not statistically significant, overall fear-potentiated startle tended to be stronger in the high IU/TA group relative to the low group (Stimulus X Group $F(1,43) = 3.81$, $p = .07$, $\eta^2_p = 0.07$). In line with the startle data, we also found strong CS discrimination in the SCRs during the first half of extinction training (trials 01–10) in the high IU/TA group, but not in the low group (Fig. 3 C and D). The effect was completely reduced in the second half (trials 11–20), resulting again in a significant quadratic trend (Group X Stimulus X Block $F(1,36) = 5.99$, $p < .05$, $\eta^2_p = 0.14$; Fig. 2 A).

### 3.2.3. BOLD fMRI

Summary statistics for significantly activated clusters within the different phases of extinction training (trials 01–20), groups (high/low IU/TA), and contrasts (CS+ > CS− and CS− < CS+) are reported in Tables S1-S4 and visualized in Figs. S4-S7. Across all blocks of extinction training (trials 01–20) and both stimuli we found a significant main effect Group within, among other significant clusters, bilateral thalamus and bilateral putamen (Fig. 4 A; Table 2). Post-hoc t-tests showed that these group differences were mainly driven by stronger activation in the low as compared to the high IU/TA group. Interestingly, when comparing the groups for the CS+ only, we additionally observed a significant cluster in the dACC/dmPFC, which was again stronger for the low IU/TA group. When looking at the CS- only, we found a significant group difference in the bilateral hippocampus for the low > high IU/TA group. Across the whole extinction training phase (trials 01–20) CS discrimination (CS+ vs. CS−) did not differ significantly between groups. However, we did find a significant interaction of Stimulus X Group in the first block of extinction training in the left vlPFC, precuneus, bilateral angular and precentral gyrus (Fig. 4 B; Table 2), which was absent in the other extinction training blocks. The pattern in the left vlPFC suggests that the high IU/TA group shows more pronounced CS discrimination due to a strong deactivation towards the CS− as compared with the low IU/TA group (Fig. 3 C). Moreover, we observed time-sensitive changes in CS discrimination from early (trials 01–05) to late extinction training (trials 16–20) that differed between groups (interaction Stimulus X Block X Group) within the left superior parietal lobule (no. voxels: 389; MNI coordinates peak voxel: $x = -26$, $y = -68$, $z = 52$; $F = 13.67$; $p$-value FWE-corrected: peak-level: $0.964$, cluster-level: $0.031$). When lowering the cluster threshold, a small cluster in the left dlPFC emerged (no. voxels: 137; MNI coordinates peak voxel: $x = -26$, $y = 30$, $z = 42$; $F = 11.88$; $p$-value FWE-corrected: peak-level: $0.999$, cluster-level: $0.645$). However, due to the lowered threshold this result should be interpreted very carefully.

### 3.3. Reinstatement test and re-extinction (day 2)

#### 3.3.1. Subjective ratings

As compared to the last rating during extinction training, valence was rated as more unpleasant after reinstatement US trials for both CS+ and CS− in the physiological sub-sample (Block $F(1,45) = 4.18$, $p < .05$, $\eta^2_p = 0.09$) irrespective of IU/TA group. In contrast, in the MRI sub-sample the low IU/TA group, but not the high group showed an increase of overall negative valence ratings (Block X Group $F(1,91) = 4.17$, $p < .05$, $\eta^2_p = 0.04$). In general, CS+ and CS− continued to differentiate strongly ($F(1,45) = 4.18$, $p < .05$, $\eta^2_p = 0.09$; $F(1,91) = 20.81$, $p < .001$, $\eta^2_p = 0.19$). For arousal ratings, we found an increase, which was comparable between both stimuli in the physiological sub-sample.
found a significant main effect Group across both stimuli within bilateral thalamus, bilateral aINS/inferior frontal gyrus, right putamen, supplementary motor cortex and right fusiform gyrus (Fig. 5 A; Table 3). Post-hoc t-tests showed that group differences were driven by stronger activation in the low as compared to the high IU/TA group, mainly towards the CS+ (Table 3). Additionally, we tested for potential group differences related to a reinstatement effect (interaction Group X Stimulus X Block [pre vs. post reinstatement]). No cluster survived our corrected significance threshold, but when exploring smaller effects by lowering the cluster threshold, we found an effect in the left dlPFC (k = 106, p < .005). During the subsequent re-extinction block, we found a significant interaction of Stimulus X Group within the left dlPFC extending to the dmPFC, and the supplementary motor cortex (Fig. 5 B; Table 3). The interaction resulted from increased activation to the CS+ as compared to the CS- in the high IU/TA group, while the low IU/TA group showed stronger activation to the CS+ as compared to the CS- (Fig. 5 C).

4. Discussion

The overarching aim of the present study was to systematically investigate how dispositional negativity affects common outcomes of fear and extinction learning (subjective ratings of negative valence, arousal and US expectancy, SCR, FPS, and neural activation) in a newly developed research paradigm with varying states of uncertainty of threat and safety. Here, dispositional negativity was assessed by two closely-related sub-constructs namely self-reported intolerance of uncertainty (IU) and trait anxiety (TA). Our results indicate that effects of dispositional negativity depend on the experimental phase and learning outcome suggesting a differentiated influence on specific processes during threat and safety processing. Mainly, we found that high levels of IU and TA went along with stronger CS+/CS- discrimination on SCR and startle responding during early extinction training, an experimental phase characterized by a high level of uncertainty regarding threat and safety. Moreover, we observed differences in extinction learning between high and low IU/TA groups in several regions relevant for threat processing and regulation (e.g., thalamus, putamen, lateral prefrontal cortex). Furthermore, we found increased dispositional negativity related to increased neural activation towards the safety cue (CS-) compared to the learned fear cue (CS+) within relevant regions of the fear network (e.g., thalamus, anterior insula) during the reinstatement test phase.

During fear acquisition training, we found robust threat learning already during early trials, indexed by higher responses to the threat-related cue (CS+) than the safety-related cue (CS-) across all outcomes. Importantly, prior to acquisition training we explicitly instructed the contingency between CS+ and US. Precise contingency instructions lead to low levels of uncertainty about the probabilistic structure of the experiment as the expectations of the participants about possible aversive events are already confirmed during early trials (Morriss et al., 2021a). Previous research suggested that higher dispositional negativity, mostly indicated by IU scores, affects CS+/CS- differentiation under conditions of high levels but not moderate levels of threat uncertainty (Chin et al., 2016). Hence, it is not surprising that we did not find general differences between low and high levels of IU/TA during the fear acquisition training as designed in our study. This result is also in line with the majority of previous findings (Mertens and Morriss, 2021; Morriss, 2019; Morriss et al., 2015, 2020; Morriss et al., 2019a; Morriss et al., 2016b; Morriss et al., 2019b; Morriss et al., 2021b; Morriss and van Reekum, 2019; Wake et al., 2021; Wake et al., 2020). However, we did observe a trend for a small effect of increasedSCRs to the threat cue relative to the safety cue in the high IU/TA group during intermediate fear acquisition. This might point to subtle effects of dispositional negativity on the temporal processes of fear learning, even under conditions of low uncertainty. Furthermore, although the level of uncertainty about CS-US associations was relatively low, uncertainty regarding the reinforcement rate or the occurrence of the startle probes.
Importantly, discrimination returned during intermediate extinction training and decreased again during later phases. Thus, low dispositional negativity might be associated with high adaptivity to changing contextual conditions as would have been indexed by a decreasing CS discrimination between CS+ and CS− during the whole extinction phase. These participants did not fully adapt to changing contextual conditions as would have been indexed by a decreasing CS discrimination during the course of extinction training, suggesting high uncertainty about threat and safety changed during intermediate extinction training 24 h after the fear acquisition training. Replicating previous results (Hollandt et al., 2020) all participants reported higher US expectancies, higher arousal ratings, and more negative valence ratings for the CS− but not for the CS+ during the first trials of extinction training as compared to the last trials of fear acquisition training indicating facilitated fear responding to the safety cue. Because no explicit instructions about US presentations and the contingency to the CSs were given in this phase, a subtle change in the learning context might have provoked a brief short-term switch to a "better-safe-than-sorry" processing strategy (Van den Bergh et al., 2021), resulting in an adaptive increase of responding to the now uncertain safety cue. After confirmatory trials of no CS−/US pairing, the safety cue was processed as a function of extinction learning. In contrast, participants with high dispositional negativity showed a robust physiological negativity might be associated with high adaptivity to changing states of uncertainty. High uncertainty during initial extinction learning might have provoked a brief short-term switch to a "better-safe-than-sorry" processing strategy (Van den Bergh et al., 2021), resulting in an adaptive increase of responding to the now uncertain safety cue. After confirmatory trials of no CS−/US pairing, the safety cue was processed as a signal for safety again. Finally, conditioned responses also decreased for the CS+ as a function of extinction learning. Importantly, discrimination returned during intermediate extinction training and decreased again during later phases. Thus, low dispositional negativity might be associated with high adaptivity to changing states of uncertainty. High uncertainty during initial extinction learning might have provoked a brief short-term switch to a "better-safe-than-sorry" processing strategy (Van den Bergh et al., 2021), resulting in an adaptive increase of responding to the now uncertain safety cue. After confirmatory trials of no CS−/US pairing, the safety cue was processed as a signal for safety again. Finally, conditioned responses also decreased for the CS+ as a function of extinction learning. In contrast, participants with high dispositional negativity showed a robust physiological discrimination between CS+ and CS− during the whole extinction training phase. These participants did not fully adapt to changing contextual conditions as would have been indexed by a decreasing CS discrimination during the course of extinction training, suggesting high uncertainty about threat and safety changed during intermediate extinction training 24 h after the fear acquisition training. Replicating previous results (Hollandt et al., 2020) all participants reported higher US expectancies, higher arousal ratings, and more negative valence ratings for the CS− but not for the CS+ during the first trials of extinction training as compared to the last trials of fear acquisition training indicating facilitated fear responding to the safety cue. Because no explicit instructions about US presentations and the contingency to the CSs were given in this phase, a subtle change in the learning context might have provoked a brief short-term switch to a "better-safe-than-sorry" processing strategy (Van den Bergh et al., 2021), resulting in an adaptive increase of responding to the now uncertain safety cue. After confirmatory trials of no CS−/US pairing, the safety cue was processed as a signal for safety again. Finally, conditioned responses also decreased for the CS+ as a function of extinction learning. In contrast, participants with high dispositional negativity showed a robust physiological discrimination between CS+ and CS− during the whole extinction training phase. These participants did not fully adapt to changing contextual conditions as would have been indexed by a decreasing CS discrimination during the course of extinction training, suggesting high uncertainty about threat and safety changed during intermediate extinction training 24 h after the fear acquisition training. Replicating previous results (Hollandt et al., 2020) all participants reported higher US expectancies, higher arousal ratings, and more negative valence ratings for the CS− but not for the CS+ during the first trials of extinction training as compared to the last trials of fear acquisition training indicating facilitated fear responding to the safety cue. Because no explicit instructions about US presentations and the contingency to the CSs were given in this phase, a subtle change in the learning context might have provoked a brief short-term switch to a "better-safe-than-sorry" processing strategy (Van den Bergh et al., 2021), resulting in an adaptive increase of responding to the now uncertain safety cue. After confirmatory trials of no CS−/US pairing, the safety cue was processed as a signal for safety again. Finally, conditioned responses also decreased for the CS+ as a function of extinction learning. In contrast, participants with high dispositional negativity showed a robust physiological discrimination between CS+ and CS− during the whole extinction training phase. These participants did not fully adapt to changing contextual conditions as would have been indexed by a decreasing CS discrimination during the course of extinction training, suggesting high uncertainty about threat and safety changed during intermediate extinction training 24 h after the fear acquisition training. Replicating previous results (Hollandt et al., 2020) all participants reported higher US expectancies, higher arousal ratings, and more negative valence ratings for the CS− but not for the CS+ during the first trials of extinction training as compared to the last trials of fear acquisition training indicating facilitated fear responding to the safety cue. Because no explicit instructions about US presentations and the contingency to the CSs were given in this phase, a subtle change in the learning context might have provoked a brief short-term switch to a "better-safe-than-sorry" processing strategy (Van den Bergh et al., 2021), resulting in an adaptive increase of responding to the now uncertain safety cue. After confirmatory trials of no CS−/US pairing, the safety cue was processed as a signal for safety again. Finally, conditioned responses also decreased for the CS+ as a function of extinction learning. In contrast, participants with high dispositional negativity showed a robust physiological discrimination between CS+ and CS− during the whole extinction training phase. These participants did not fully adapt to changing contextual conditions as would have been indexed by a decreasing CS discrimination during the course of extinction training, suggesting
reduced extinction learning. Instead, they continued to respond according to learned threat associations, which were probably strongly established during the instructed fear acquisition training. High dispositional negativity has been suggested previously to result from a stagnating error-reduction process, which results from a generalized but non-adaptive better-safe-than-sorry strategy (Van den Bergh et al., 2021). In our study, this might have resulted in a failure to reassess risk as environmental conditions and uncertainty changed. Indeed, high levels of IU and (trait) anxiety have earlier been linked to cognitive rigidity and poor adaptive behavior (Biasi et al., 2015; Fergus and Rowatt, 2014; Kesby et al., 2019; Morris and Mansell, 2018; Schultz and Seareleman, 2002), especially in social contexts (Lamba et al., 2020).

The observed dynamics between dispositional negativity and threat/safety processing during changing states of uncertainty might also explain previously observed heterogeneous findings. Previous research has shown, that even subtle differences in methodology significantly affect risk perception and threat evaluations, e.g., the degree of similarity between conditioned threat and safety cues, the degree of (temporal) coincidence between conditioned cues and unconditioned stimuli, reinforcement rates, or explicit information about the experimental contingencies (Lonsdorf et al., 2017; Lonsdorf and Richter, 2017). Again, this suggests that fear and especially extinction learning highly depend on the degree of perceived uncertainty. Depending on individual participants these uncertainties can change from trial to trial and hence, extensive analyses approaches, e.g., computational models of single-trial events and uncertainties, should be considered in future research. The situation regarding the impact of trait anxiety on CS evoked psychophysiological responses during extinction training is slightly different. Previous studies failed to demonstrate a consistent link between trait anxiety and CS evoked SCRs. Unlike most of the previous findings reporting effects of IU and trait anxiety mainly during late extinction training (Bauer et al., 2020; Morriss, 2019; Morriss et al., 2015; Morriss et al., 2016a; Morriss et al., 2016b; Morriss and van Reekum, 2019; Wake et al., 2020, 2021), our results indicate increased CS discrimination in FPS and SCRs during early extinction training for the high IU/TA group. At least for FPS, increased CS discrimination continues across the extinction phase. A possible reason could be that previous studies rather focused on immediate instead of delayed extinction training, which leads to pronounced fear recall at the beginning of extinction training. Again, computational models of trial-by-trial events could benefit the interpretation of these effects.

In contrast to the physiological data, group effects were not observed for subjective ratings. This is in line with previous results (Morriss et al., 2021b), although some studies reported a poor discrimination between CS+ vs. CS- in US expectancy ratings in individuals with high IU (Morriss et al., 2019c), and a trend for increased fear ratings of the CS+
vs. the CS during delayed extinction training (Klingelhöfer-Jens et al., 2021). However, a recent study identified a specific trajectory for fear ratings of the CSs, with high scores during un instructed phases of fear acquisition and extinction training, but substantially reduced scores during instructed phases (Leen et al., 2021). Thus, further research is needed to substantiate the notion stemming from the present study, that during instructed phases (Leen et al., 2021). Thus, further research is needed. In the current analyses we mainly found an overall decreased activation to both CSs in the bilateral thalamus and putamen in the group of high IU/TA. The thalamus is a cognitive hub in the basolateral limbic circuit and passes sensory information to the amygdala and mPFC (see Levy and Schiller, 2021 for review). In the animal model, a suppression of thalamus activation leads to impaired extinction learning, indicating a central role of the thalamus not only in the expression of conditioned fear, but also in the acquisition and consolidation of extinction memory (Lee et al., 2019; Lee and Shin, 2016). Accordingly, changes in the functioning of this structure in individuals with high IU/TA, might have mediated the effects of extinction training in the current study. Furthermore, neural activation in the thalamus has previously been associated with the modulation of uncertainty (Kosciessa et al., 2021). Our additional analyses (see supplement) provide evidence that IU over TA drives the observed group differences during early, but not during late extinction training. Again, this corresponds to the changing state of uncertainty with relatively high uncertainty during early, and lower uncertainty during late extinction learning. Although not typically considered as part of the fear network, the striatum and the putamen have been discussed to play a role in anxiety (Iago et al., 2017) and stress-related disorders (Homan et al., 2019). In their recent review Levy and Schiller (2021) showed that the striatum is not only associated with avoidance, but also with decision making and uncertainty. Similarly, Justin Kim et al. (2017) provided evidence that grey matter volume in the bilateral putamen is correlated with IU scores. Our findings about

### Table 3

Group comparisons (low vs. high IU/TA) of fMRI BOLD activation during reinstatement test phase and re-extinction phase. All contrasts were assessed at $p < .005$ uncorrected with a cluster threshold of $k = 175$.

<table>
<thead>
<tr>
<th>Region</th>
<th>Hemisphere</th>
<th>No. voxels</th>
<th>MNI coordinates peak voxel</th>
<th>F-/T-value</th>
<th>$p$-value FWE-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main effect group (reinstatement test phase; trials 21–25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusiform gyrus + lingual gyrus</td>
<td>R/L</td>
<td>221</td>
<td>18, –74, 30</td>
<td>19.18</td>
<td>0.264, 0.244</td>
</tr>
<tr>
<td>Anterior insula + inferior frontal gyrus</td>
<td>L</td>
<td>602</td>
<td>–42, 30</td>
<td>18.94</td>
<td>0.289, 0.003</td>
</tr>
<tr>
<td>Thalamus</td>
<td>R</td>
<td>417</td>
<td>14, 20</td>
<td>15.50</td>
<td>0.782, 0.022</td>
</tr>
<tr>
<td>Thalamus</td>
<td>L</td>
<td>400</td>
<td>–14, 12</td>
<td>15.39</td>
<td>0.798, 0.027</td>
</tr>
<tr>
<td>Putamen</td>
<td>R</td>
<td>256</td>
<td>26, 14, 10</td>
<td>15.09</td>
<td>0.838, 0.157</td>
</tr>
<tr>
<td>Anterior insula + inferior frontal gyrus</td>
<td>R</td>
<td>394</td>
<td>38, 36, 2</td>
<td>14.93</td>
<td>0.857, 0.029</td>
</tr>
<tr>
<td>Middle cingulate gyrus + supplementary motor cortex</td>
<td>R</td>
<td>326</td>
<td>0, 10, 54</td>
<td>14.03</td>
<td>0.943, 0.066</td>
</tr>
</tbody>
</table>

Post-hoc t-test: CS+: low IU/TA ≥ high IU/TA (reinstatement test phase; trials 21–25)

<table>
<thead>
<tr>
<th>Hemisphere</th>
<th>No. voxels</th>
<th>MNI coordinates peak voxel</th>
<th>F-/T-value</th>
<th>$p$-value FWE-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thalamus</td>
<td>L</td>
<td>1939</td>
<td>–10, –10, 20</td>
<td>4.44</td>
</tr>
<tr>
<td>Thalamus</td>
<td>R</td>
<td>2462</td>
<td>12, –12, 20</td>
<td>4.40</td>
</tr>
<tr>
<td>Precentral gyrus</td>
<td>R</td>
<td>461</td>
<td>52, 6, 40</td>
<td>4.38</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>L</td>
<td>373</td>
<td>–46, –54, 14</td>
<td>4.34</td>
</tr>
<tr>
<td>Supramarginal gyrus</td>
<td>L</td>
<td>407</td>
<td>54, 36, 26</td>
<td>4.34</td>
</tr>
<tr>
<td>Middle cingulate gyrus</td>
<td>R</td>
<td>1008</td>
<td>0, 0, 32</td>
<td>3.75</td>
</tr>
<tr>
<td>Supplementary motor cortex</td>
<td>L</td>
<td>209</td>
<td>0, –14, 48</td>
<td>3.66</td>
</tr>
<tr>
<td>Dorsolateral prefrontal cortex</td>
<td>L</td>
<td>205</td>
<td>–24, 38, 32</td>
<td>3.51</td>
</tr>
</tbody>
</table>

Post-hoc t-test: CS+: low IU/TA ≥ high IU/TA (reinstatement test phase; trials 21–25)

<table>
<thead>
<tr>
<th>Hemisphere</th>
<th>No. voxels</th>
<th>MNI coordinates peak voxel</th>
<th>F-/T-value</th>
<th>$p$-value FWE-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putamen + frontal operculum</td>
<td>L</td>
<td>288</td>
<td>–32, –18, 6</td>
<td>3.84</td>
</tr>
</tbody>
</table>

Post-hoc t-test: CS+: low IU/TA ≤ high IU/TA (reinstatement test phase; trials 21–25)

No sig. activation

Interaction Stimulus X Group (re-extinction phase; trials 26–30)

<table>
<thead>
<tr>
<th>Hemisphere</th>
<th>No. voxels</th>
<th>MNI coordinates peak voxel</th>
<th>F-/T-value</th>
<th>$p$-value FWE-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorsolateral + dorsomedial prefrontal gyrus</td>
<td>L</td>
<td>462</td>
<td>–22, 46, 34</td>
<td>18.82</td>
</tr>
<tr>
<td>Supplementary motor cortex</td>
<td>L</td>
<td>303</td>
<td>–4, 22, 66</td>
<td>14.67</td>
</tr>
</tbody>
</table>

Abbreviations: CS+: conditioned stimulus that is followed by the unconditioned stimulus (US) with a reinforcement rate of 60 % (only unpaired CS+ were included); CS-: conditioned stimulus that is never followed by a US; L: left; R: right; no. voxel: number of voxels per cluster; x, y, z: MNI coordinates.
decreased activation in the putamen may reflect a misinterpretation and prediction error about the state of uncertainty during extinction training.

When separately comparing learned threat (CS+) and safety cues (CS-) between groups across the whole extinction training phase, we identified further differences in regions belonging to the extended fear network. Regarding the CS+, we found an increased activation in the dACC/dmPFC in participants with low IU/TA, and an increased activation for the CS- in the hippocampus. The dACC/dmPFC has been previously linked to the processing of threat-related stimuli and fear expression (Chavanne and Robinson, 2021; Fullana et al., 2018; Milad et al., 2007; Sehmeyer et al., 2009). In contrast, the hippocampus appears to be more closely associated with the regulation of conditioned threat responses and memory formation (Fullana et al., 2018; Levy and Schiller, 2021; Sevenster et al., 2018). At first glance, our findings seem counterintuitive. However, Sehmeyer et al., 2011 reported a reduced dACC activation during extinction in individuals with high trait anxiety, which further supports our results. As discussed earlier, individuals with high IU/TA may suffer from impaired adaptive behavior and thus reduced dACC activation and hippocampal deactivation during extinction training may reflect an impeded updating of CS-US contingencies (Browning et al., 2015). Furthermore, previous more time-sensitive analyses revealed further associations of IU/TA with ventro- and dorsolateral PFC activity, especially during early extinction training. From decision making and reward research, the lateral PFC is known to contribute to a neural representation of uncertainty (Bach et al., 2009; Huettel et al., 2006; Mohr et al., 2010; Preuschhoff et al., 2006; Tobler et al., 2007). This finding is further supported by our additional dimensional analyses (see supplement) showing that activation in the left vPFC is mainly predicted by IU over TA during early extinction training, but not during late extinction training. Thus, reduced vPFC activation to the CS+ during early extinction training in high IU/TA may reflect maladaptive behavioral responses to changes in CS-US contingencies.

Similar to the context change before extinction training, the return of fear manipulation (i.e., reinstatement) may also produce an unexpected increase of uncertainty about the predictive value of the CSs regarding the occurrence of the US, which decreases fast during re-extinction (Morriss et al., 2021b). Previous studies suggested a modulatory effect of IU on the return of fear as evident in conditioned SCRs (Dunsmoor et al., 2015; Lucas et al., 2018). However, the pattern regarding effects of trait anxiety on the return of fear appears to be less clear (Gazendam et al., 2013; Kindt et al., 2009; Kindt and Soeter, 2013; Martinez et al., 2012; Soeter and Kindt, 2010). Our data do not suggest a significant association between IU/TA and the effects of reinstatement, neither for subjective ratings nor for psychophysiological measures. Potentially, more time-sensitive analyses would be beneficial to investigate ROF, as the effects may be diminished within the blocks of five trials in our analyses. However, this would be out of the scope of the current study and should be considered in future analyses. Regarding neural activation, we found again stronger activation to the threat-related cue in a number of regions of the fear network, such as the thalamus, for low levels of IU/TA. This finding provides further evidence for impaired adaptive processing in individuals with high levels of dispositional negativity (Biasi et al., 2015; Fergus and Rowatt, 2014; Kesby et al., 2019; Lamba et al., 2020; Morris and Mansell, 2018; Schultz and Searleman, 2002).

Besides a number of strengths, the present study has to be interpreted in the light of some limitations. First, no patients with psychiatric disorders were included in the current analyses. This resulted in reduced variability in IUS and STAI-T scores, which are expected to be significantly higher in clinical samples (McEvoy et al., 2019). As compared to other studies, especially the IUS scores in the current sample seem to be lower than in other healthy samples (Buhr and Dugas, 2002). This could be one possible reason, why we found only small differences between groups. Second, we used an instructed fear acquisition training protocol, which might have reduced experienced uncertainty during acquisition training. However, the general focus of the PROTECT-AD study was on examining links between (neuro-)biological mechanisms of extinction learning and exposure-based CBT (Heinig et al., 2017; Pittig et al., 2021) and the experimental paradigm was optimized for that purpose. Future studies should consider to investigate the effects of different uncertainty manipulations on conditioned responding. Third, the decision on using faces as CSs could potentially influence learning behavior during fear conditioning depending on ethnicity and gender (Mallan et al., 2009; Mazurski et al., 1996; Navarrete et al., 2009). However, the paradigm was designed to tailor it towards translation to exposure-based therapy procedures and hence, faces represent socially more relevant stimuli, which show stable conditioning effects (Hollandt et al., 2020; Ridderbusch et al., 2021). Fourth, data-driven cluster analyses should rather be seen as an exploratory approach to derive hypotheses that need to be tested thoroughly in future studies. Fifth, although the overall sample (n = 155) was rather large compared to previous studies, the sample was subdivided into a psychophysiological and MRI sub-sample on day 2. Hence, the resulting numbers of participants per IU/TA group in each sub-sample, as well as the power for dimensional analyses were reduced.

Taken together the present study provides a systematic investigation of dispositional negativity on behavioral, psychophysiological, and neural outcome measures of fear conditioning in an experimental paradigm optimized for clinical translation. Our results demonstrate that IU/TA is mainly related to various measures of conditioned responding in experimental phases with high uncertainty, e.g., during early delayed extinction training. Our results point to elevated levels of cognitive rigidity in individuals with high IU/TA, i.e., altered learning patterns under changing environments, stemming from psychophysiological and neural measures of conditioned responding. These observations might benefit the treatment of mental disorders by furthering our understanding of the precise mechanisms underlying the effects of individual personality characteristics on extinction learning, thereby providing a tool to promote positive outcomes of exposure-based treatments. However, in future studies more time-sensitive analyses should be considered to shed light on the exact nature of these observations.

Staff members by site


Ethical standards

The study program is performed according to the Declaration of Helsinki and was approved by the Ethics Committee of Technische Universität Dresden (EK 234062014, November 14, 2014). The clinical trial in adults has been registered with NIMH Protocol Registration System (01EE1402A) and with the German Register of Clinical Studies.


Carlet, R.N., 2016a. Fear of the unknown: one fear to rule them all? J. Anxiety Disord. 41, 5–21.

Carlet, R.N., 2016b. Into the unknown: a review and synthesis of contemporary models involving uncertainty. J. Anxiety Disord. 39, 30–43.


