Intolerance of uncertainty and threat generalization: a replication and extension

Article

Accepted Version

Bauer, E. A., MacNamara, A., Sandre, A., Lonsdorf, T. B., Weinberg, A., Morriss, J. and Van Reekum, C. M. ORCID: https://orcid.org/0000-0002-1516-1101 (2020) Intolerance of uncertainty and threat generalization: a replication and extension. Psychophysiology, 57 (5). e13546. ISSN 1540-5958 doi: https://doi.org/10.1111/psyp.13546 Available at https://centaur.reading.ac.uk/89089/

www.reading.ac.uk/centaur

CentAUR

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Intolerance of uncertainty and threat generalization: A replication and extension

Elizabeth A. Bauer[1] [ID] | Annmarie MacNamara[1] [ID] | Aislinn Sandre[2] |
Tina B. Lonsdorf[3] | Anna Weinberg[2] [ID] | Jayne Morriss[4] [ID] | Carien M. van Reekum[4]

[1]Department of Psychological and Brain Sciences, Texas A&M University, College Station, TX, USA

[2]Department of Psychology, McGill University, Montreal, QC, Canada

[3]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[4]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

Correspondence
Elizabeth A. Bauer, Department of Psychological and Brain Sciences, Texas A&M University, 4235 TAMU, College Station, TX 77843, USA.
Email: eabauer@tamu.edu

Funding information
National Institute of Mental Health, Grant/Award Number: K23MH105553; Brain and Behavior Research Foundation, Grant/Award Number: 27567; Economic and Social Research Council, Grant/Award Number: ES/R01145/1

Abstract

Intolerance of uncertainty (IU) is a transdiagnostic risk factor for internalizing disorders. Prior work has found that IU may be associated with either increased reactivity to threat or, alternatively, with decreased differential responding between threat and nonthreat/safety cues (i.e., threat generalization). For example, work by Morriss, Macdonald, & van Reekum (2016) found that higher IU was associated with increased threat generalization during acquisition (using skin conductance response (SCR)), as well as less differentiation between acquisition and extinction (using subjective uneasiness ratings). Here, three labs attempted direct and conceptual replications of Morriss, Macdonald, et al. (2016). Results showed that the direct replication failed, despite being conducted at the same lab site as the original study; moreover, in contrast to Morriss, Macdonald, et al. (2016), the direct replication found that higher IU was associated with greater SCR *discrimination* between threat and safety cues (across acquisition and extinction), as well as greater differences in uneasiness ratings between acquisition and extinction. Nonetheless, in the conceptual replications, higher IU was associated with greater threat *generalization*, as well as less discrimination between acquisition and extinction, as measured using SCR. Higher IU was also associated with larger late positive potentials to threat versus safety cues during extinction—results that mirror those observed by Morriss, Macdonald, et al. (2016) using SCR. Results are discussed with regards to the challenge involved in defining a successful replication attempt, the benefits of collaborative replication and the use and reliability of multiple measures.

KEYWORDS
event-related potentials (ERPs), intolerance of uncertainty (IU), skin conductance response (SCR), threat generalization

1 | INTRODUCTION

Intolerance of uncertainty (IU), a dispositional tendency to perceive uncertainty as threatening (Carleton, 2016), is associated with the presence of and risk for internalizing disorders

Elizabeth A. Bauer and Annmarie MacNamara shared first-authorship.

(Gentes & Ruscio, 2011; Jensen, Cohen, Mennin, Fresco, & Heimberg, 2016). In the lab, higher self-reported IU has been linked to heightened reactivity to uncertain threat (for review, see Tanovic, Gee, & Joormann, 2018). Furthermore, individual differences in IU have been shown to modulate threat generalization (Morriss, Christakou, & van Reekum, 2016; Morriss, Macdonald, & van Reekum, 2016; Nelson,

Weinberg, Pawluk, Gawlowska, & Proudfit, 2015). For example, one prevailing finding emerging from this work has been that IU is associated with greater threat generalization—that is, reduced discrimination between threatening stimuli and perceptually similar safe stimuli. Yet heightened neural reactivity to threatening stimuli in the absence of uncertainty might be expected in higher IU individuals (e.g., Schienle, Köchel, Ebner, Reishofer, & Schäfer, 2010). Reconciling these conflicting conceptualizations of IU as it relates to threat processing (i.e., increased threat generalization vs. increased reactivity to threat) is essential if its role in the development of internalizing pathology is to be understood.

Conditioning paradigms consist of an unconditioned stimulus (US) paired with a conditioned stimulus (CS+) to result in a conditioned response. In some threat generalization paradigms, a CS+ (e.g., a large square) serves as a cue for an aversive US (e.g., an unpleasant noise). Generalization stimuli (GS) that resemble the CS+ to varying degrees (e.g., squares of increasing/decreasing size) are also presented but are not paired with the US. Threat acquisition training can be followed by threat extinction training, in which both the CS+ and GS are presented but the US is no longer presented. During acquisition training, reduced differential responding between the CS+ and GS is interpreted as evidence of threat generalization (see Morriss, Christakou, et al., 2016; Morriss, Macdonald, et al., 2016; Nelson et al., 2015).

To determine whether IU is associated with threat generalization as evidenced by self-reported uneasiness ratings and skin conductance response (SCR), Morriss, Macdonald, et al. (2016) used a CS+ and three perceptually graded GS that resembled the CS+ to varying degrees (GS1, most similar; GS2, somewhat similar; GS3, least similar). Results showed that across participants, self-reported uneasiness was parametrically graded according to the perceptual similarity between the CS+ and the GS; additionally, SCR was greater for the CS+ versus GS2 and GS3, as well as the GS1 versus GS3, although there were no significant differences between the stimuli that were most perceptually similar (i.e., CS+ vs. GS1; GS2 vs. GS3). Whereas lower IU participants showed larger SCRs for the CS+ and GS1 versus the GS2, higher IU participants failed to distinguish between any of the stimuli presented during acquisition training, suggesting greater threat generalization as measured by SCR. Individuals with higher IU also showed greater discrimination between the CS+ and the GS during extinction training as measured by SCR, and less reduction in uneasiness ratings between training phases, suggesting delayed learning of cue pairings (as higher IU participants did not differentiate between stimuli until extinction training). Notably, these results were evident controlling for individual differences in trait anxiety and worry.

In addition to SCR, other measures, such as event-related potentials (ERPs), can be used to assess threat processing. For example, the late positive potential (LPP) is a centroparietal ERP component beginning around 300 ms after stimulus onset that is larger for emotional compared to neutral stimuli (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Hajcak, MacNamara, & Olvet, 2010). Using the LPP and a threat generalization paradigm, Nelson and colleagues (2015) found that the LPP was larger for a CS+ compared to the GS, and that the magnitude of LPP did not differ between GS across participants. In addition, participants with higher prospective IU, a component of IU, showed *greater* discrimination between the CS+ and the GS compared to participants with lower prospective IU (Nelson et al., 2015). These results differ from the Morriss, Macdonald, et al. (2016) findings, where SCR distinguished between the GS and high IU was associated with reduced discrimination/increased threat generalization.

Here, we attempted a three-lab direct and conceptual replication of the Morriss, Macdonald, et al. (2016) original results. Recent work has suggested that replication should aim to test the boundaries and generalizability of prior theories and confirm the robustness of theoretical predictions outside of the original study parameters (Nosek & Errington, 2019). As such, we included two conceptual replications in the current study to test if the theoretical basis behind the Morriss, Macdonald, et al. (2016) findings would hold. University of Reading, the lab from Morriss, Macdonald, et al. (2016), used the same paradigm parameters as in the original study (direct replication). Conceptual replication and extension was attempted at McGill University (conceptual replication 1) and Texas A&M University (conceptual replication 2) using uneasiness ratings and SCR but also ERPs; paradigm parameters were modified to facilitate the elicitation of ERPs. We hypothesized that at the direct and conceptual level, uneasiness ratings, SCR, and ERPs in acquisition and extinction would be parametrically graded to the threat and safety cues such that responding would be greatest for the threat cue, followed by cues that were more perceptually similar to the threat cue (Morriss, Macdonald, et al., 2016). Despite evidence to the contrary (Nelson et al., 2015), we reasoned that we would observe parametric gradation of the LPP, because similar findings have been observed using several other measures (e.g., startle, Hajcak et al., 2009; insula reactivity, Greenberg, Carlson, Cha, Hajcak, & Mujica-Parodi, 2013; SCR, Morriss, Macdonald, et al., 2016), and because our paradigm contained fewer and simpler GS than that used by Nelson and colleagues (2015; three GS of increasing *or* decreasing size vs. six GS of increasing *and* decreasing size), which should simplify distinctions between stimuli. Furthermore, we hypothesized that graded responding to stimuli would be more evident in acquisition training, in which the CS+ predicted an aversive outcome, compared to extinction training, in which all stimuli were equivalent/none of the stimuli predicted an aversive outcome (Morriss, Macdonald, et al., 2016). In line with Morriss, Macdonald, et al. (2016), we hypothesized that higher IU would be

associated with (a) greater threat generalization during acquisition training and (b) increased responding to threat cues during extinction training. We expected to observe these individual differences using SCR and the LPP, but not necessarily in the ratings (Morriss, Macdonald, et al., 2016). Finally, we also hypothesized that associations between threat generalization and IU would hold when controlling for trait anxiety and worry (Morriss, Macdonald, et al., 2016).

2 | METHOD

All data are open and available on the Open Science Framework (https://osf.io/vuj8r/). We report all conditions, measures, manipulations, and data exclusions.[1]

2.1 | Direct replication

2.1.1 | Participants

Results from the original study had suggested an *N* of at least 66 to achieve 0.8 power (Morriss, Macdonald, et al., 2016). Participants in the direct replication were an unselected community sample recruited from the University of Reading and consisted of 84 participants (67 female; *M* age = 22.7 years, range = 18–48 years; 61.25% White European, 6.25% Middle Eastern, 10% Asian, 3.75% mixed race, 18.75% not specified). Participants received financial compensation for their participation.

2.1.2 | Self-report measures

The Intolerance of Uncertainty Scale (IUS; Dugas, Buhr, & Ladouceur, 2004) is a 27-item measure assessing response to uncertain or ambiguous situations. Participants respond on a 5-point scale from 1 (*not at all characteristic of me*) to 5 (*entirely characteristic of me*). Higher scores on the IUS indicate greater IU. Internal consistency for the IUS was excellent within the direct replication's sample, $\alpha = .94$.

The State-Trait Anxiety Inventory—Trait version (STAI-X2; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is a 20-item measure assessing trait anxiety. Participants respond on a 4-point scale from 1 (*almost never*) to 4 (*almost always*), with higher scores indicating greater trait anxiety. Internal consistency for the STAI-X2 was excellent within the direct replication's sample, $\alpha = .93$.

The Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990) is a 16-item measure

assessing worry. Participants respond on a 5-point scale from 1 (*not at all typical of me*) to 5 (*very typical of me*), with higher scores on the PSWQ indicating greater levels of worry. Internal consistency for the PSWQ was good within the direct replication's sample, $\alpha = .83$.

2.1.3 | Stimuli

The CS+ and GS consisted of the same stimuli used by Morriss, Macdonald, et al. (2016): four yellow squares of increasing/ decreasing size. The CS+ was the largest or smallest square (counterbalanced across participants) and the GS were squares of increasing or decreasing size (GS1, GS2, GS3), with the GS1 being closest in size to the CS+ and the GS3 being the most different. As in Morriss, Macdonald, et al. (2016), the US was a female scream (sound 277, International Affective Digitized Sound battery-2; Bradley & Lang, 2007). The scream was shortened to 1,000 ms in length and amplified by 15 dB, resulting in a 90-dB sound, as verified by audiometer. At the end of each block, participants used a 1 (*not at all*) to 9 (*extremely*) scale to rate how "uneasy" they felt about each visual stimulus.

2.1.4 | Procedure

The same procedure was used as in Morriss, Macdonald, et al. (2016). Participants completed a series of questionnaires followed by a perceptual discrimination task to ensure that they could distinguish between stimuli. Participants then completed acquisition and extinction training from the Morriss, Macdonald, et al. (2016) study. They were instructed to keep their eyes on the screen at all times and they were told that during the task, they would be asked to rate how uneasy they felt about the stimuli. Visual stimuli were presented on a black background on a 13.3 in (33.8 cm) monitor; participants were seated approximately 60 cm from the screen. The unconditioned auditory stimulus was played over headphones. Stimuli were presented using E-Prime 2.0 (Psychology Software Tools Ltd, Pittsburgh, PA).

The task consisted of acquisition and extinction training phases involving the presentation of the CS+ and GS. During acquisition training, the CS+ was paired with the aversive scream 50% of the time, while GS were always presented alone. In extinction training, none of the stimuli were paired with the scream. The direct replication used the timing from the original study (Morriss, Macdonald, et al., 2016), with the CS+ and GS presented for 1,500 ms and a variable inter-trial interval (ITI) ranging between 4,700 and 6,500 ms. The US/scream played during the last 1,000 ms of visual stimulus presentation and coterminated with the visual stimuli, as in the original study (Morriss, Macdonald, et al., 2016). The ITI consisted of a white fixation cross presented on a black background.

---

[1]Analyses for other ERP components, the P2, P3, and the stimulus preceding negativity (SPN) can be found in Supplemental Materials.

As in Morriss, Macdonald, et al. (2016), acquisition training consisted of four blocks with 9 trials in each block, resulting in a total of 36 trials (12 CS+ [6 paired and 6 unpaired], 8 GS1, 8 GS2, 8 GS3); extinction training consisted of four blocks with 10 trials in each block, resulting in a total of 40 trials (10 CS+ unpaired, 10 GS1, 10 GS2, 10 GS3). Trials were pseudo-randomized into an order with the constraint that the same stimulus could not be presented more than three times in a row (as in the original study). At the end of each block, participants viewed the visual stimuli (CS+, GS1, GS2, GS3), presented in random order, and were asked to rate how "uneasy" they felt about each square.

2.1.5 | Uneasiness ratings

As in Morriss, Macdonald, et al. (2016), uneasiness ratings were averaged separately for each stimulus and training phase (acquisition, extinction), yielding eight averages per participant. The final sample size for ratings analyses was 80 participants (64 female; M age = 22.7 years, range = 18–48 years).

2.1.6 | SCL recording and data reduction

The same SCL recording and data reduction procedure was used as in Morriss, Macdonald, et al. (2016). Skin conductance level (SCL) was recorded using ADinstruments hardware and was digitized through a 16-bit A/D converter and sampled at a frequency of 1,000 Hz. A constant-voltage AC excitation of 22 mV$_{rms}$ at 75 Hz was passed through the electrodes, which were connected to an ML116 GSR Amp, and converted to DC before being digitized and stored. Electrodes were placed on the distal phalanges of the index and middle fingers on the participant's nondominant hand. The amplitude of each SCR was scored as the difference between the onset (i.e., the trough preceding the highest peak) and the offset (i.e., maximum deflection before the signal decreased) of the response. The direct replication's lab used an ADinstruments macro within LabChart software to detect onsets and offsets with a derivative function. As in the original study (Morriss, Macdonald, et al., 2016), only SCR onsets that occurred within 500 to 4,000 ms following stimulus presentation were included in analyses; SCR was scored when there was an increase in SCL exceeding 0.03 microSiemens (µS) and trials without valid SCRs were scored as zero and were included in analyses at all sites. SCR magnitudes were square root transformed to reduce skewness prior to z-scoring within subjects to control for interindividual differences in skin conductance responsiveness. Paired CS+ trials were excluded from the SCR analyses.

To assess whether individuals learned to discriminate between the stimuli that predicted threat or safety, difference score metrics were calculated by subtracting average GS3 SCR magnitude from the CS+ SCR magnitude, separately for acquisition and extinction training. In line with the original manuscript (Morriss, Macdonald, et al., 2016), individuals who failed to show greater SCRs (i.e., any positive difference) for the CS+ versus the GS3 in either acquisition or extinction training were identified as nondiscriminators and were excluded from SCR analyses (n = 14). Discriminators did not differ from nondiscriminators on IU, trait anxiety, or worry at any lab site, all ps > .50. In addition, participants who did not have SCRs on at least 10% of the total number of trials used in SCR analyses (i.e., greater than 90% of trials, excluding paired CS+ trials, were scored as 0) were identified as nonresponders and were also excluded (n = 3). Final sample size for SCR was 63 participants (49 female; M age = 22.5 years, range = 18–48 years).

2.2 | Conceptual replication 1

2.2.1 | Participants

Participants consisted of unselected undergraduate students from McGill University. The sample consisted of 96 participants (72 female; M age = 21.6 years, range = 18–34 years; 50% Caucasian, 15.6% Chinese, 4.2% South Asian, 6.3% Arab/West Asian, 5.2% Hispanic, 18.6% other). Participants received financial compensation or undergraduate course credit.

2.2.2 | Self-report measures

The same self-report measures were used as described in the direct replication. Internal consistency for each measure was as follows: IUS, α = .92; STAI-X2, α = .92; PSWQ, α = .93.

2.2.3 | Stimuli

The same stimuli were used as described in the direct replication.

2.2.4 | Procedure

The direct replication procedure was used, with the following modifications:

Visual stimuli were presented on a black background on a 19 in (48.3 cm) monitor. Stimuli were presented using Presentation software (Neurobehavioral Systems Inc., Berkeley, CA). Because reviewers for the Morriss, Macdonald, et al. (2016)

study had commented that the timing used in the original paradigm was somewhat short for analysis of SCR (stimuli presented for 1,500 ms, ITI between 4,700 and 6,500 ms), we adapted the task timing for conceptual replication 1, such that the CS+ and GS were presented for 6,000 ms and the ITI varied between 6,000 ms and 8,000 ms. As in Morriss, Macdonald, et al. (2016), the US/scream played during the last 1,000 ms of visual stimulus presentation and coterminated with the visual stimuli. Further, because reviewers of Morriss, Macdonald, et al. (2016) had commented on the fixed trial order used in that study, trial order was random within the following constraints: in the acquisition training phase, the first trial was always a paired CS+ and each block had 3 CS+, 2 GS1, 2 GS2 and 2 GS3 presentations, and in the extinction training phase, no more than three trials of the same type were presented in a row.

2.2.5 | Uneasiness ratings

The same uneasiness rating averages were used as described in the direct replication. The final sample size for ratings analyses was 89 participants (67 female; M age = 21.6 years, range = 18–34 years).

2.2.6 | SCL recording and data reduction

The same SCL recording and data reduction procedures were used as in the direct replication, with the following modifications:

SCL was recorded using the galvanic skin response add-on module for the ActiCHamp amplifier (Brain Products GmbH, Gilching, Germany) which were digitized through a 24-bit A/D converter and sampled at a frequency of 1,000 Hz. A constant-voltage DC excitation of 0.5 V was passed through the electrodes. The peak detection tool in Brain Vision Analyzer 2 was used (Brain Products GmbH, Gilching, Germany). Data were also reviewed manually (e.g., to ensure that the maximum occurred after the trough). Data were segmented in Brain Vision Analyzer, beginning at stimulus onset and lasting for 9,000 ms. In keeping with Morriss, Macdonald, et al. (2016), SCR onsets were identified from 500 to 4,000 ms following stimulus onset. However, given the longer stimulus presentation duration and different ITI, SCR offsets were identified following SCR onset and anytime up to 8,000 ms following stimulus onset.[2]

---

[2]The permissible time range for SCR offsets was decided based on the timings used in both conceptual replications, which allowed the same processing parameters to be used for both conceptual replications 1 and 2. Conceptual replication 2 had an ITI of 1,000 to 3,000 ms; 6,000 ms stimulus presentation plus a 1,000 ms ITI (minimum) and an additional 1,000 ms for an SCR to be elicited from the next stimulus yielded an 8,000 ms window during which SCR offset could be observed without interference from the next trial.

As in the original study and direct replication, individuals who failed to show greater SCRs (i.e., any positive difference) for the CS+ versus the GS3 in either acquisition or extinction training were identified as nondiscriminators and were excluded from SCR analyses ($n = 11$). No participants from conceptual replication 1 were identified as nonresponders. Additionally, SCR was not recorded from 42 participants due to equipment failures. The final sample size for SCR analyses in conceptual replication 1 was 43 participants (30 female; M age = 21.5 years, range = 18–33 years).

2.2.7 | EEG recording and data reduction

Continuous EEG recordings were collected using an ActiCap and the ActiCHamp amplifier system (Brain Products GmbH, Gilching, Germany). Thirty-two electrode sites were used based on the 10/20 system and impedances were kept below 30 kΩ. The electrooculogram (EOG) was recorded from four facial electrodes: two that were placed approximately 1 cm above and below the right eye, forming a bipolar channel to measure vertical eye movement and blinks and two that were placed approximately 1 cm beyond the outer edges of each eye, forming a bipolar channel to measure horizontal eye movements. The EEG data were digitized at 24-bit resolution and a sampling rate of 1,000 Hz.

EEG data were processed offline using Brain Vision Analyzer 2 software (Brain Products GmbH, Gilching, Germany). The signal from each electrode was re-referenced to the average of the left and right mastoids (TP9/TP10) and band-pass filtered with high-pass and low-pass filters of 0.01 and 30 Hz, respectively. Data were segmented for each trial beginning 200 ms prior to stimulus onset and continuing for 5,000 ms (to avoid the period of time associated with the 1,000-ms scream), and averages were created separately for each condition. Baseline correction for each trial was performed using the 200 ms prior to stimulus onset. Eye blink and ocular corrections used the method developed by Miller, Gratton, and Yee (1988). Artifact analysis was used to identify a voltage step of more than 50.0 μV between sample points, a voltage difference of 300.0 μV within a trial, and a maximum voltage difference of less than 0.50 μV within 100-ms intervals. Trials were also inspected visually for any remaining artifacts, and data from individual channels containing artifacts were rejected on a trial-to-trial basis.

As specified a priori, the LPP was scored by averaging amplitudes at CP1, CP2, Cz, and Pz from 400 to 1,000 ms, 1,000 to 2,500 ms, and 2,500 to 5,000 ms following stimulus onset. More than one time window for the LPP was used based on prior research suggesting that the LPP is comprised of multiple positivities rather than one sustained positivity, with early and late portions of the component reflecting distinct

processes (see Foti, Hajcak, & Dien, 2009). Participants were excluded from EEG analyses if they had artifacts on 50% or more trials ($n = 23$). The final sample size for EEG analyses in conceptual replication 1 was 70 participants (55 female; M age = 22.0 years, range = 18–34 years).

2.3 | Conceptual replication 2

2.3.1 | Participants

Participants consisted of unselected undergraduate students from Texas A&M University. The sample consisted of 74 participants (47 female; M age = 19.4 years, range = 18–22 years; 66.2% white, 4.1% African American, 1.4% American Indian/Alaskan Native, 9.5% Asian, 2.7% Native Hawaiian/Pacific Islander, 16.4% mixed race or other; 38.4% Hispanic/Latino). Participants received undergraduate course credit for their participation.

2.3.2 | Self-report measures

The same self-report measures were used as described in the direct replication. Internal consistency for each measure was as follows: IUS, $\alpha = .95$; STAI-X2, $\alpha = .92$; PSWQ, $\alpha = .95$.

2.3.3 | Stimuli

The same stimuli were used as described in the direct replication.

2.3.4 | Procedure

The direct replication procedure was used, with the following modifications:

Visual stimuli were presented on a black background on a 21.5 in (54.6 cm) monitor. Stimuli were presented using Presentation software (Neurobehavioral Systems Inc., Berkeley, CA). For conceptual replication 2, we used timing optimized for ERPs, with the CS+ and GS presented for 6,000 ms and a variable ITI ranging between 1,000 ms and 3,000 ms. As in Morriss, Macdonald, et al. (2016), the US/scream played during the last 1,000 ms of visual stimulus presentation and coterminated with the visual stimuli.

As in conceptual replication 2, trial order was random within the following constraints: in the acquisition training phase, the first trial was always a paired CS+ and each block had 3 CS+, 2 GS1, 2 GS2, and 2 GS3 presentations, and in the extinction training phase, no more than three trials of the same type were presented in a row across blocks.

2.3.5 | Uneasiness ratings

The same uneasiness rating averages were used as in the direct replication and conceptual replication 1. The final sample size for ratings analyses was 74 participants (47 female; M age = 19.4 years, range = 18–22 years).

2.3.6 | SCL recording and data reduction

The same SCL recording and data reduction procedures were used as in conceptual replication 1.

Individuals identified as nondiscriminators were excluded from SCR analyses ($n = 18$). No participants from conceptual replication 2 were identified as nonresponders. The final sample size for SCR analyses was 56 participants (36 female; M age = 19.4 years, range = 18–22 years).

2.3.7 | EEG recording and data reduction

The same EEG recording and data reduction procedures were used as in conceptual replication 1. The final sample size for EEG analyses in conceptual replication 2 was 67 participants (42 female; M age = 19.5 years, range = 18–22 years).

2.4 | Data analyses

Following analyses used in the original study (Morriss, Macdonald, et al., 2016), separate multilevel models were constructed for ratings, SCR, and ERPs by entering stimulus (CS+, GS1, GS2, GS3) and phase (acquisition, extinction) as Level 1 factors and individual subjects as a Level 2 factor, with IU, PSWQ, and STAI-X2 entered as continuous individual difference predictor variables at Level 2. A diagonal covariance matrix was used for Level 1, and random effects included a random intercept for each individual subject, where a variance components covariance structure was used. Fixed effects included stimulus and phase. A maximum likelihood estimator was also used. Also, in line with Morriss, Macdonald, et al. (2016), significant interactions with IU were followed up using pairwise comparisons, performed separately for participants 1 SD above the mean for IU and 1 SD below the mean on the regression line for IU.[3] These follow-up analyses were estimated from the multilevel model of the entire sample, not unlike performing a

---

[3]The following values were entered into the MLM for Low IU and High IU. All variables were mean centered: Morriss, Macdonald et al. (2016), $\mp$ 16.69; direct replication uneasiness ratings, $\mp$ 18.87; direct replication SCR, $\mp$ 19.01; conceptual replication 1 uneasiness ratings, $\mp$ 17.99; conceptual replication 1 SCR, $\mp$ 14.78; conceptual replication 1 ERPs, $\mp$ 17.21; conceptual replication 2 uneasiness ratings, $\mp$ 19.39; conceptual replication 2 SCR, $\mp$ 20.57; conceptual replication 2 ERPs, $\mp$ 19.86.

TABLE 1 Mean self-report scores [95% confidence intervals] for each replication attempt

	IUS *M* [95% CI]	STAIX-2 *M* [95% CI]	PSWQ *M* [95% CI]
Direct replication	62.34 [58.14, 66.54]	42.74 [40.38, 45.10]	48.65 [45.79, 51.52]
Conceptual 1	64.48 [60.92, 68.04]	43.54 [41.17, 45.91]	50.96 [48.18, 53.74]
Conceptual 2	58.74 [54.25, 63.24]	41.22 [38.72, 43.71]	51.32 [47.87, 54.77]

Note: Morriss, Macdonald, et al. (2016) participants had the following average scores on the self-report measures: IUS *M* = 60.77, *SD* = 16.69; STAIX-2 *M* = 42.88, *SD* = 9.30; PSWQ *M* = 48.16, *SD* = 12.19.

Abbreviations: IUS, Intolerance of Uncertainty Scale; PSWQ, Penn State Worry Questionnaire; STAIX-2, State-Trait Anxiety Inventory—Trait version.

TABLE 2 SCR and LPP even–odd reliability coefficients in acquisition and extinction for all stimuli

Measure		Acquisition				Extinction			
		CS+ (*r*)	GS1 (*r*)	GS2 (*r*)	GS3 (*r*)	CS+ (*r*)	GS1 (*r*)	GS2 (*r*)	GS3 (*r*)
SCR	Direct replication	−0.46	0.14	−0.23	0.34	0.34	0.31	0.04	0.13
	Conceptual 1	0.17	0.36	−0.12	0.66*	−0.23	0.19	−0.12	0.52*
	Conceptual 2	0.55*	−0.16	0.35	0.20	0.10	0.36	−0.20	0.23
LPP (400–1000 ms)	Conceptual 1	0.10	0.06	0.23	0.58*	0.34	0.51*	0.45*	0.35
	Conceptual 2	0.38	0.20	0.18	0.31	0.26	−0.23	0.18	0.42*
LPP (1000–2500 ms)	Conceptual 1	0.11	−0.07	−0.07	0.40*	0.37	0.25	0.18	−0.27
	Conceptual 2	0.22	0.33	0.42*	0.01	0.20	0.01	0.11	0.17
LPP (2500–5000 ms)	Conceptual 1	−0.06	0.09	−0.20	0.58*	0.53*	0.35	0.57*	−0.05
	Conceptual 2	−0.05	0.19	0.14	0.14	−0.12	0.12	−0.18	−0.35

*p < .05.

simple slopes analysis in a multiple regression analysis. Given the large number of variables and conditions involved, we only report overall experimental effects (i.e., main effects of stimulus and phase and their interaction) and interactions that involved IU. Multilevel models were constructed separately at each site as our intent was to determine whether effects could be replicated using each of three, somewhat different approaches (rather than to examine effects of site or interactions with site). The internal consistency of psychophysiological measures was assessed using even–odd reliability, in which correlations were performed between averages created separately for even and odd trials. The Spearman–Brown formula was used to correct these correlations (Nunnally, 1978). All analyses were conducted at Texas A&M University using SPSS statistical software version 24.0 (IBM, Armonk, NY).[4]

3 | RESULTS

Table 1 presents average scores for self-report measures for each lab site. Table 2 presents even–odd reliability coefficients for SCR and ERPs for each stimulus in acquisition and extinction training. Table 3 presents uneasiness ratings, Table

---

[4]We also conducted these same analyses for prospective and inhibitory IU, separately. Results can be found in Supplemental Materials.

4 presents SCR magnitudes, and Table 5 presents ERP amplitudes. Estimated marginal means are presented separately for each stimulus in acquisition and extinction training and for each lab site for uneasiness ratings, SCR magnitudes, and ERP amplitudes. Finally, Table 6 presents a summary of results across lab sites including Morriss, Macdonald, et al. (2016).

3.1 | Direct replication

3.1.1 | Ratings

Figure 1 depicts average uneasiness ratings in acquisition and extinction training for the direct replication, shown separately for participants lower and higher in IU.

Significant main effects of phase, $F(1, 353.55) = 114.66$, $p < .001$ (acquisition > extinction) and stimulus, $F(3, 316.65) = 125.10$, $p < .001$ (CS+ > GS1 > GS2 > GS3, all ps < .001), were qualified by an interaction between phase X stimulus, $F(3, 316.65) = 5.50$, $p = .001$. Follow-up analyses showed that in acquisition training, participants' uneasiness ratings were highest for the CS+ and decreased incrementally with each GS (CS+ > GS1 > GS2 > GS3, all ps < .001). Likewise, in extinction training, participants' uneasiness ratings were also highest for the CS+ and decreased with each GS (CS+ > GS1 > GS2 > GS3, all ps < .02). Furthermore, participants' uneasiness ratings were

TABLE 3 Mean uneasiness ratings [95% confidence intervals] in acquisition and extinction for all stimuli

	Acquisition					Extinction				
	CS+ M [95% CI]	GS1 M [95% CI]	GS2 M [95% CI]	GS3 M [95% CI]		CS+ M [95% CI]	GS1 M [95% CI]	GS2 M [95% CI]	GS3 M [95% CI]	
Direct replication	5.31 [4.83, 5.79]	4.10 [3.70, 4.50]	3.06 [2.71, 3.41]	2.48 [2.12, 2.85]		4.10 [3.65, 4.55]	2.85 [2.49, 3.22]	2.30 [1.96, 2.64]	2.01 [1.66, 2.37]	
Conceptual 1	4.45 [3.75, 5.16]	3.42 [2.88, 3.97]	2.77 [2.34, 3.21]	2.47 [2.02, 2.93]		3.00 [2.43, 3.57]	2.48 [2.04, 2.91]	2.09 [1.71, 2.47]	1.98 [1.60, 2.37]	
Conceptual 2	4.69 [4.05, 5.33]	3.59 [3.05, 4.14]	2.48 [2.02, 2.94]	2.19 [1.74, 2.64]		3.50 [2.84, 4.16]	2.60 [2.13, 3.08]	2.17 [1.74, 2.60]	1.87 [1.42, 2.31]	

Note: Participants made uneasiness ratings on a scale from 1 to 9, with 9 indicating greater uneasiness.

TABLE 4 Mean SCR [95% confidence intervals] in acquisition and extinction for all stimuli

	Acquisition M [95% CI] μS					Extinction M [95% CI] μS				
	CS+	GS1	GS2	GS3		CS+	GS1	GS2	GS3	
Direct replication	0.26 [0.16, 0.35]	0.18 [0.09, 0.27]	−0.03 [−0.12, 0.06]	−0.05 [−0.13, 0.04]		0.13 [0.04, 0.22]	0.01 [−0.08, 0.10]	−0.12 [−0.20, −0.05]	−0.26 [−0.33, −0.19]	
Conceptual 1	0.31 [0.18, 0.44]	0.05 [−0.06, 0.15]	0.08 [−0.01, 0.17]	0.00 [−0.10, 0.11]		−0.01 [−0.09, 0.08]	−0.11 [−0.20, −0.02]	−0.04 [−0.14, 0.06]	−0.16 [−0.27, −0.06]	
Conceptual 2	0.19 [0.06, 0.32]	0.15 [0.05, 0.25]	0.06 [−0.05, 0.16]	0.00 [−0.10, 0.10]		0.04 [−0.04, 0.12]	−0.12 [−0.22, −0.03]	−0.07 [−0.16, 0.02]	−0.13 [−0.21, −0.05]	

Note: SCR values are square root transformed and z-scored. Means presented are estimated marginal means (i.e., taking into account the individual difference variables) and confidence intervals.

TABLE 5 Mean LPP amplitudes [95% confidence intervals] in acquisition and extinction for all stimuli

		Acquisition M [95% CI] μV				Extinction M [95% CI] μV			
		CS+	GS1	GS2	GS3	CS+	GS1	GS2	GS3
LPP (400–1000 ms)	Conceptual 1	1.63 [0.89, 2.37]	0.67 [−0.47, 1.81]	1.43 [0.56, 2.40]	1.49 [0.42, 2.56]	2.21 [1.25, 3.17]	−0.14 [−1.14, 0.87]	0.89 [−0.14, 1.92]	0.79 [−0.31, 1.89]
	Conceptual 2	1.75 [0.87, 2.63]	1.16 [0.10, 2.22]	1.07 [0.21, 1.92]	1.46 [0.43, 2.48]	3.44 [2.67, 4.22]	0.71 [−0.06, 1.48]	1.11 [0.32, 1.89]	1.05 [0.24, 1.86]
LPP (1000–2500 ms)	Conceptual 1	−0.47 [−1.87, 1.04]	−1.45 [−1.55, 2.05]	−0.20 [−1.35, 1.83]	−0.75 [−1.32, 1.98]	−0.05 [−1.04, 1.87]	−1.70 [−2.05, 1.55]	−0.45 [−1.83, 1.35]	−1.09 [−1.98, 1.32]
	Conceptual 2	−0.60 [−1.60, 0.41]	−0.30 [−1.89, 1.30]	−0.72 [−1.92, 0.49]	−0.82 [−2.21, 0.57]	1.09 [0.30, 2.15]	−1.05 [−2.11, 0.02]	0.34 [−0.64, 1.31]	−0.73 [−1.62, 0.16]
LPP (2500–5000 ms)	Conceptual 1	0.30 [−1.37, 1.98]	−1.05 [−2.98, 0.89]	1.38 [−0.44, 3.19]	1.34 [−0.95, 3.62]	−0.24 [−2.07, 1.60]	−0.81 [−2.68, 1.06]	0.57 [−1.42, 2.56]	−0.80 [−2.44, 0.84]
	Conceptual 2	0.09 [−1.16, 1.35]	1.14 [−0.82, 3.10]	0.65 [−0.93, 2.22]	0.00 [−1.84, 1.83]	1.99 [0.62, 3.36]	−0.52 [−1.86, 0.82]	0.96 [−0.18, 2.10]	−0.11 [−1.21, 1.00]

higher for each stimulus in acquisition training compared to extinction training, all $ps < .002$. There was also a significant interaction between phase X IU, $F(1, 353.55) = 5.72$, $p = .02$. While participants lower and higher in IU both made greater uneasiness ratings in acquisition training compared to extinction training, $ps < .001$, the reduction in uneasiness ratings between phases was larger for higher IU participants compared to lower IU participants, contrary to results from Morriss, Macdonald, et al. (2016). No other effects reached significance at the omnibus level, all $ps > .51$.

3.1.2 | SCR

Figure 2 depicts average SCR magnitude for each stimulus across phases in the direct replication, shown separately for participants lower and higher in IU.

There were significant main effects of phase, $F(1, 486.37) = 25.71$, $p < .001$ (acquisition > extinction) and stimulus, $F(3, 242.14) = 27.60$, $p < .001$ (CS+ > GS1 > GS2, GS3, $ps < .03$; GS2 > GS3, $p = .05$). The phase X stimulus interaction observed by Morriss, Macdonald, et al. (2016) did not reach significance, $F(3, 242.14) = 0.82$, $p = .49$; however, there was a significant interaction between stimulus X IU, $F(3, 242.14) = 3.19$, $p = .02$. Follow-up tests showed that participants with lower IU showed larger SCRs for the CS+ compared to the GS3 and GS1 compared to the GS3, $ps < .02$, whereas SCRs to other stimuli did not differ, $ps > .14$. Participants with higher IU showed greater SCRs for the CS+ compared to all other stimuli and for the GS1 compared to the GS2 and GS3 (CS+ > GS1, GS2, GS3, $ps < .009$; GS1 > GS2, GS3, $ps < .004$; GS2 > GS3, $p = .20$). The phase X stimulus X IU interaction observed by Morriss, Macdonald, et al. (2016) did not reach significance, $F(3, 242.14) = 1.19$, $p = .31$. The interaction between phase X IU also did not reach significance, $F(1, 486.37) = 1.12$, $p = .29$.

3.2 | Conceptual replication 1

3.2.1 | Ratings

Significant main effects of phase, $F(1, 357.39) = 124.45$, $p < .001$ (acquisition > extinction) and stimulus, $F(3, 274.67) = 45.44$, $p < .001$ (CS+ > GS1 > GS2 > GS3, all $ps < .02$) were qualified by an interaction between phase X stimulus, $F(3, 274.67) = 3.66$, $p = .01$ Follow-up analyses showed that in acquisition training, participants' uneasiness ratings were highest for the CS+ and decreased incrementally with each GS (CS+ > GS1, GS2, GS3, all $ps < .003$; GS1 > GS2, GS3, all $ps < .001$; GS2 vs. GS3, $p = .11$). Uneasiness ratings were also highest for the CS+ and decreased incrementally with

TABLE 6 Summary of results across replication attempts and from Morriss, Macdonald, et al. (2016)

		Phase	Stimulus	Phase X Stimulus	Phase X IU	Stimulus X IU	Phase X Stimulus X IU
Uneasiness Ratings	Morriss, Macdonald, et al. (2016)	$F(1, 169.44) = 74.10^{***}$ Acq > Ext	$F(3, 168.02) = 77.26^{***}$ CS+>GS1 > GS2>GS3	$F(3, 168.02) = 4.79^{**}$ Acq: CS+>GS1 > GS2>GS3 Ext CS+>GS1 > GS2>GS3	$F(1, 169.44) = 4.55^{*}$ Low IU: Acq > Ext> High IU: Acq > Ext	ns	ns
	Direct replication	$F(1, 353.55) = 114.66^{***}$ Acq > Ext	$F(3, 316.65) = 125.10^{***}$ CS+>GS1 > GS2>GS3	$F(3, 316.65) = 5.50^{**}$ Acq: CS+>GS1 > GS2>GS3 Ext CS+>GS1 > GS2>GS3	$F(1, 353.55) = 5.72^{*}$ Low IU: Acq > Ext < High IU: Acq > Ext	$F(3, 316.65) = 0.77$, ns	$F(3, 316.65) = 0.06$, ns
	Conceptual 1	$F(1, 357.39) = 124.45^{***}$ Acq > Ext	$F(3, 279.21) = 68.01^{***}$ CS+>GS1 > GS2>GS3	$F(3, 274.67) = 3.66^{*}$ Acq: CS+>GS1,GS2,GS3; GS1 > GS2,GS3 Ext CS+>GS1 > GS2>GS3	$F(1, 357.39) = 0.01$, ns	$F(3, 274.67) = 0.76$, ns	$F(3, 274.67) = 0.05$, ns
	Conceptual 2	$F(1, 308.46) = 61.40^{***}$ Acq > Ext	$F(3, 279.21) = 68.01^{***}$ CS+>GS1 > GS2>GS3	$F(3, 279.21) = 3.60^{*}$ Acq: CS+>GS1 > GS2>GS3; Ext CS+>GS1 > GS2>GS3	$F(1, 308.46) = 2.12$, ns	$F(3, 279.21) = 0.33$, ns	$F(3, 279.21) = 0.05$, ns
SCR	Morriss, Macdonald, et al. (2016)	$F(1, 308.35) = 16.17^{***}$ Acq > Ext	$F(3, 165.23) = 8.31^{***}$ CS+>GS1,GS2,GS3 GS1 > GS3	$F(3, 165.23) = 6.28^{***}$ Acq: ns Ext CS+>GS1,GS2,GS3 GS1 > GS3	ns	ns	$F(3, 165.23) = 4.24^{**}$ Low IU: Acq CS+, GS1 > GS2; Ext CS+,GS1,GS2 > GS3 High IU: Acq ns Ext CS+>GS1,GS2,GS3; GS1 > GS3
	Direct replication	$F(1, 486.37) = 25.71^{***}$ Acq > Ext	$F(3, 242.14) = 27.60^{***}$ CS+>GS1 > GS2,GS3	$F(3, 242.14) = 0.82$, ns	$F(1, 486.37) = 1.12$, ns	$F(3, 242.14) = 3.19^{*}$ Low IU: CS+>GS3; GS1 > GS3 High IU: CS+>GS1,GS2,GS3; GS1 > GS2,GS3	$F(3, 242.14) = 1.19$, ns
	Conceptual 1	$F(1, 318.56) = 28.63^{***}$ Acq > Ext	$F(3, 152.91) = 7.06^{***}$ CS+>GS1,GS2,GS3; GS2 > GS3	$F(3, 152.91) = 1.33$, ns	$F(1, 318.56) = 7.08^{**}$ Low IU: Acq > Ext High IU: ns	$F(3, 152.91) = 0.33$, ns	$F(3, 152.91) = 2.62$, ns [#] Low IU: Acq CS+>GS1,GS2; Ext GS2 > GS3 High IU: Acq CS+,GS2 > GS3
	Conceptual 2	$F(1, 402.71) = 23.65^{***}$ Acq > Ext	$F(3, 204.54) = 4.25^{**}$ CS+>GS2,GS3	$F(3, 204.54) = 1.01$, ns	$F(1, 402.71) = 0.77$, ns	$F(3, 204.54) = 3.28^{*}$ Low IU: CS+>GS1,GS2,GS3 High IU: GS2 > GS3	$F(3, 204.54) = 1.72$, ns

(Continues)

TABLE 6 (Continued)

		Phase	Stimulus	Phase X Stimulus	Phase X IU	Stimulus X IU	Phase X Stimulus X IU
LPP (400–1000ms)	Conceptual 1	$F(1, 458.56) = 1.39$, ns	$F(3, 221.59) = 5.36**$ CS+>GS1	$F(3, 221.59) = 1.38$, ns	$F(1, 458.56) = 0.07$, ns	$F(3, 221.59) = 0.88$, ns	$F(3, 221.59) = 1.87$, ns
	Conceptual 2	$F(1, 434.59) = 0.58$, ns	$F(3, 198.20) = 7.73***$ CS+>GS1,GS2,GS3	$F(3, 198.20) = 3.28*$ Acq: ns Ext: CS+>GS1,GS2,GS3	$F(1, 434.59) = 0.10$, ns	$F(3, 198.20) = 1.28$, ns	$F(3, 198.20) = 0.70$, ns
LPP (1000–2500 ms)	Conceptual 1	$F(1, 476.24) = 0.06$, ns	$F(3, 224.09) = 2.09$, ns	$F(3, 224.09) = 0.21$, ns	$F(1, 476.24) = 0.04$, ns	$F(3, 224.09) = 1.64$, ns	$F(3, 224.09) = 9.94$, ns
	Conceptual 2	$F(1, 396.12) = 1.75$, ns	$F(3, 199.94) = 1.58$, ns	$F(3, 199.94) = 1.80$, ns	$F(1, 396.12) = 0.03$, ns	$F(3, 199.94) = 1.23$, ns	$F(3, 199.94) = 3.40*$ Low IU: Ext GS3 > GS1 High IU: Ext CS+,GS2 > GS3
LPP (2500–5000ms)	Conceptual 1	$F(1, 467.90) = 1.78$, ns	$F(3, 215.68) = 1.64$, ns	$F(3, 215.68) = 0.61$, ns	$F(1, 467.90) = 0.03$, ns	$F(3, 215.68) = 1.31$, ns	$F(3, 215.68) = 0.66$, ns
	Conceptual 2	$F(1, 397.73) = 0.05$, ns	$F(3, 195.37) = 1.04$, ns	$F(3, 195.37) = 2.14$, ns	$F(1, 397.73) = 0.77$, ns	$F(3, 195.37) = 0.49$, ns	$F(3, 195.37) = 1.87$, ns

Note: ns $p > .05$.

Abbreviation: Acq, acquisition; Ext, extinction.

$***p < .001$; $**p < .01$; $*p < .05$; $^{#}p = .05$.

each GS in extinction training (CS+ > GS1 > GS2 > GS3, all ps < .01) and uneasiness ratings for each stimulus were higher in acquisition training than in extinction training, all ps < .001. The interaction between phase X IU observed by Morriss, Macdonald, et al. (2016) did not reach significance, $F(1, 357.39) = 0.01$, $p = .94$. No other effects reached significance at the omnibus level, all ps > .51.

3.2.2 | SCR

Figure 3 depicts average SCR magnitude in acquisition and extinction training in conceptual replication 1, shown separately for participants lower and higher in IU.

There were significant main effects of phase, $F(1, 318.56) = 28.63$, $p < .001$ (acquisition > extinction) and stimulus, $F(3, 152.91) = 7.06$, $p < .001$ (CS+ > GS1, GS2, GS3, ps < .02; GS2 > GS3, $p = .04$; all other ps > .27). The interaction between phase X stimulus observed by Morriss, Macdonald, et al. (2016) did not reach significance, $F(3, 152.91) = 1.33$, $p = .27$. In addition, a significant interaction was observed between phase X IU, $F(1, 318.56) = 7.08$, $p = .008$, such that participants with lower IU showed larger SCRs in acquisition training compared to extinction training, $p < .001$, whereas no effect of phase was observed for participants with higher IU, $p = .17$. The interaction between phase X stimulus X IU originally observed by Morriss, Macdonald, et al. (2016) did not reach significance, $F(3, 152.91) = 2.62$, $p = .05$. Nonetheless, because of the p value and because we had hypothesized this three-way interaction and it aligned with Morriss, Macdonald, et al. (2016), we chose to follow it up. Follow-up tests showed that participants with lower IU showed larger SCRs for the CS+ than the GS1 and GS2 in acquisition training, (CS+ > GS1, GS2, ps < .02; all other ps > .13). For participants with higher IU, SCRs were larger in acquisition training for the CS+ and GS2 compared to the GS3 (CS+ > GS3, GS2 > GS3, ps < .02; all other ps > .06). In extinction training, participants with lower IU showed larger SCR for the GS2 compared to the GS3, $p = .03$; all other ps > .15. Participants with higher IU showed numerically larger SCR magnitude for the CS+ than the GS1 in extinction, although this difference did not reach significance, $p = .05$; all other ps > .11. The interaction between stimulus X IU observed for the direct replication and conceptual replication 2 also did not reach significance, $F(3, 152.91) = 0.33$, $p = .81$.

3.2.3 | LPP (400–1000 ms)

There was a main effect of stimulus, $F(3, 221.59) = 5.36$, $p = .001$, with greater LPP amplitude for the CS+ than the GS1, $p < .001$. The interaction between phase X stimulus did not reach
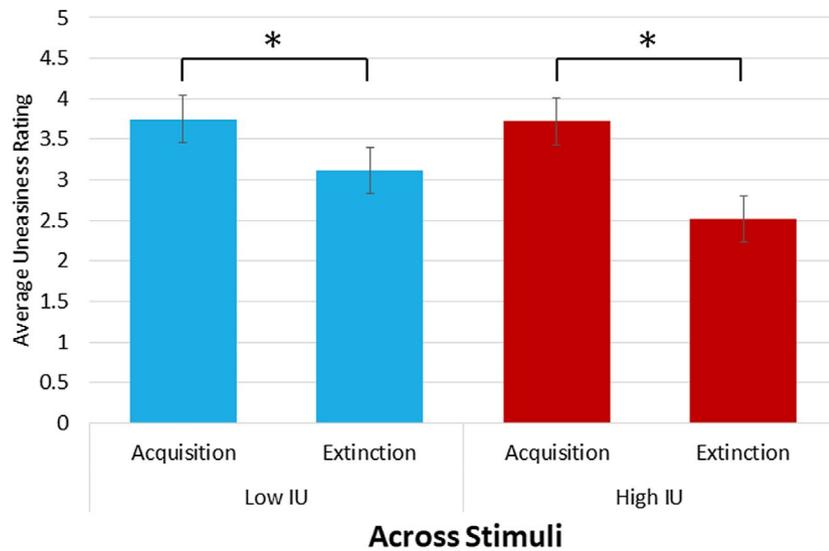
FIGURE 1 Mean uneasiness ratings (from the direct replication), shown separately for acquisition and extinction training (across stimuli), for participants with Low IU (1 *SD* below mean; left) and High IU (1 *SD* above mean; right). Error bars represent standard error of the mean. Omnibus analyses used continuous IU scores; "Low" and "High" IU groups were only created to follow-up and illustrate significant interactions. * = $p < .001$
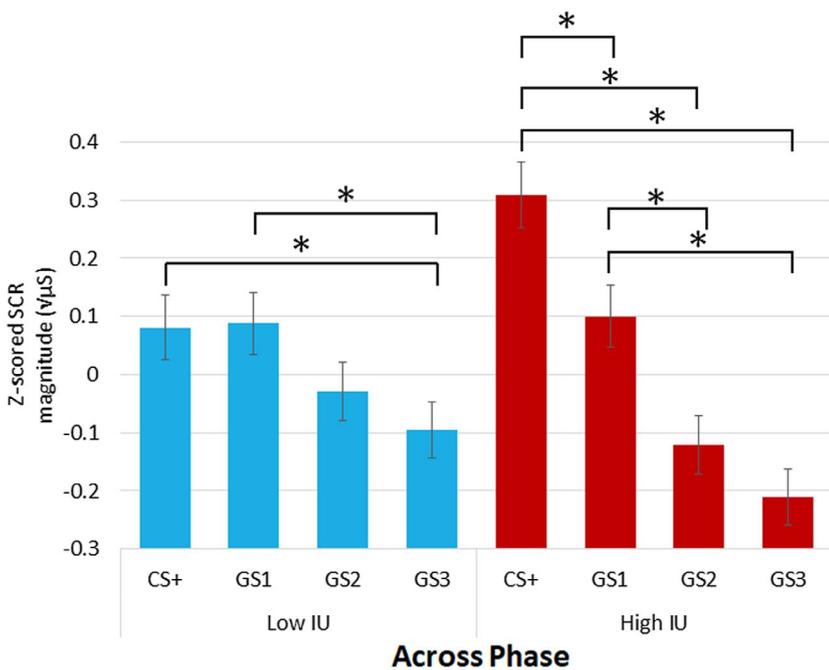

FIGURE 2 Mean SCR magnitude (from the direct replication), shown separately for each stimulus (across phase), for participants with Low IU (1 *SD* below mean; left) and High IU (1 *SD* above mean; right). Error bars represent standard error of the mean. Omnibus analyses used continuous IU scores; "Low" and "High" IU groups were only created to follow-up and illustrate significant interactions. CS+ = conditioned stimulus; GS1, GS2, GS3 = generalization stimuli. * = $p < .02$

significance, $F(3, 221.59) = 1.38$, $p = .25$. No other effects reached significance at the omnibus level, all other $ps > .13$.

3.2.4 | LPP (1000–2500 ms)

No effects reached significance, all $ps > .10$, including the interaction between phase X stimulus X IU, $F(3, 224.09) = 9.94$, $p = .42$.

3.2.5 | LPP (2500–5000 ms)

No effects reached significance, all $ps > .18$.

3.3 | Conceptual replication 2

3.3.1 | Ratings

Significant effects of phase, $F(1, 308.46) = 61.40$, $p < .001$ (acquisition > extinction) and stimulus, $F(3, 279.21) = 68.01$, $p < .001$ (CS+ > GS1 > GS2 > GS3, all $ps < .001$) were qualified by an interaction between phase X stimulus, $F(3, 279.21) = 3.60$, $p = .01$. Follow-up tests showed that in acquisition training, uneasiness was highest for the CS+ and decreased incrementally with each GS (CS+ > GS1 > GS2 > GS3, all $ps < .02$). In extinction training, participants' uneasiness ratings also decreased linearly from the CS+ to the GS3 (CS+ > GS1 > GS2 > GS3, all ps

< .003). Furthermore, participants rated their uneasiness as higher for each stimulus in acquisition training compared to extinction training, all $ps < .002$. The interaction between phase X IU observed by Morriss, Macdonald, et al. (2016) did not reach significance, $F(1, 308.46) = 2.12, p = .15$. No other effects reached significance at the omnibus level, all $ps > .80$.

3.3.2 | SCR

Figure 4 depicts average SCR magnitude for each stimulus across phases in conceptual replication 2, shown separately for participants lower and higher in IU.

There were significant main effects of phase, $F(1, 402.71) = 23.65, p < .001$ (acquisition > extinction) and stimulus, $F(3, 204.54) = 4.25, p = .006$ (CS+ > GS2, GS3, $ps < .03$; all other $ps > .05$). The interaction between phase X stimulus observed by Morriss, Macdonald, et al. (2016) was not significant, $F(3, 204.54) = 1.01, p = .39$, but there was a significant interaction between stimulus X IU, $F(3, 204.54) = 3.28, p = .02$. Follow-up tests showed that participants with lower IU showed larger SCRs for the CS+ compared to all other stimuli (CS+ > GS1, GS2, GS3, $ps < .02$; all other $ps > .15$), whereas participants with higher IU only showed larger SCRs for the GS2 compared to the GS3, $p = .01$; all other $ps > .10$. The interaction between phase X stimulus X IU observed by Morriss, Macdonald, et al. (2016) also did

FIGURE 3 Mean SCR magnitude (from conceptual replication 1), shown separately for acquisition and extinction training (across stimuli), for participants with Low IU (1 *SD* below mean; left) and High IU (1 *SD* above mean; right). Error bars represent standard error of the mean. Omnibus analyses used continuous IU scores; "Low" and "High" IU groups were only created to follow-up and illustrate significant interactions. * = $p < .001$
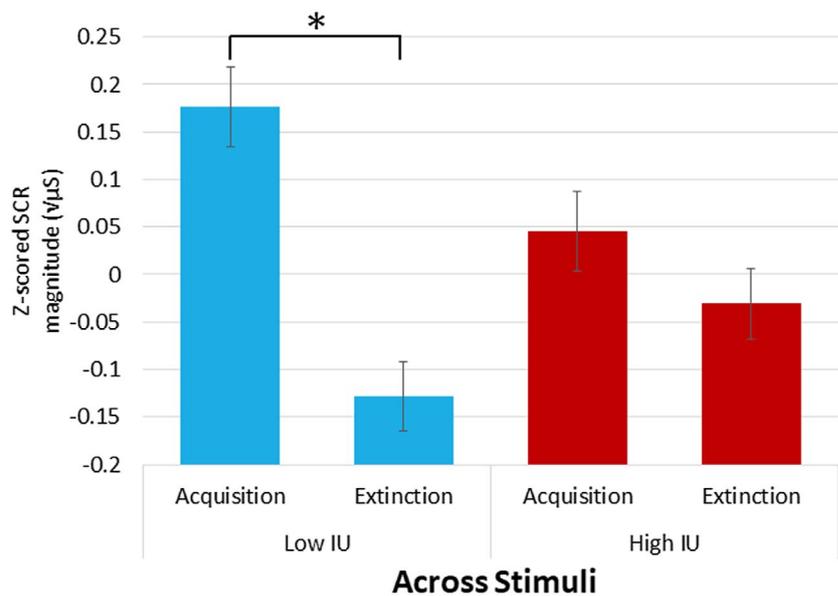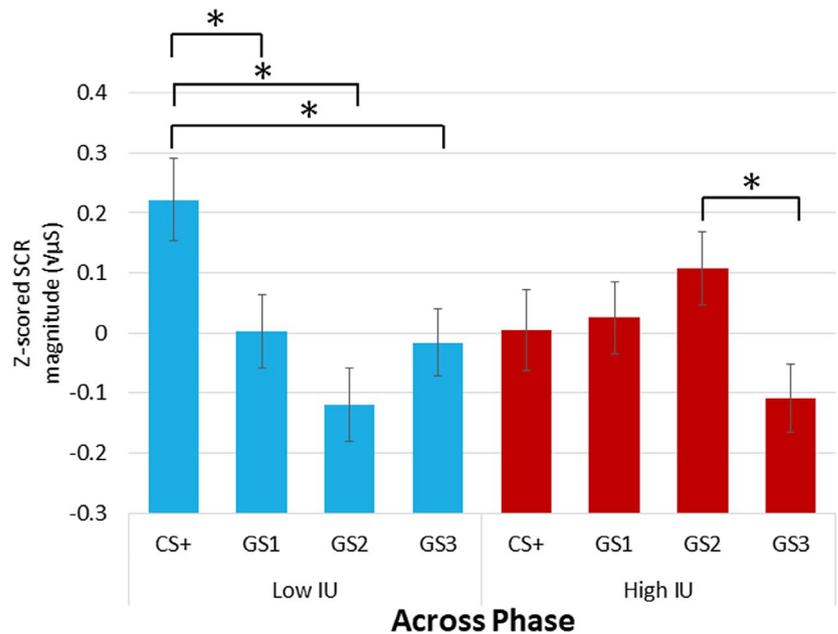

FIGURE 4 Mean SCR magnitude (from conceptual replication 2), shown separately for each stimulus (across phase), for participants with Low IU (1 *SD* below mean; left) and High IU (1 *SD* above mean; right). Error bars represent standard error of the mean. Omnibus analyses used continuous IU scores; "Low" and "High" IU groups were only created to follow-up and illustrate significant interactions. CS+ = conditioned stimulus; GS1, GS2, GS3 = generalization stimuli. * = $p < .02$

not reach significance, $F(3, 204.54) = 1.72$, $p = .16$. The interaction between phase X IU observed for conceptual replication 1 did not reach significance, $F(1, 402.71) = 0.77$, $p = .38$. No other effects reached significance at the omnibus level, all ps > .16.

3.3.3 | LPP (400–1000 ms)

A significant main effect of stimulus, $F(3, 198.20) = 7.73$, $p < .001$ (CS+ > GS1, GS2, GS3, all ps < .002) was qualified by an interaction between phase X stimulus, $F(3, 198.20) = 3.28$, $p = .02$. Follow-up tests showed that in acquisition training, LPP amplitude did not differ between any stimuli, all ps > .23. However, in extinction training, LPP amplitude was greater for the CS+ than all GS, and GS did not differ from each other (CS+ > GS1, GS2, GS3, all ps < .001; all other ps > .42). No other effects reached significance at the omnibus level, all other ps > .28.

3.3.4 | LPP (1000–2500 ms)

Figure 5 depicts grand average waveforms for stimuli in acquisition and extinction training in conceptual replication 2 at the centroparietal pooling where the LPP was scored, shown separately for participants with lower and higher IU, with headmaps depicting the distribution of amplitude differences for the CS+ minus the GS3, from 1,000 to 2,500 ms following stimulus onset.

Statistical analyses revealed a significant interaction of phase X stimulus X IU, $F(3, 199.94) = 3.40$, $p = .02$. Follow-up tests showed that for participants with lower IU, the LPP in extinction training was larger for the GS3 than the GS1, $p = .04$; all other ps > .09. For participants with higher IU, the LPP in extinction training was larger for the CS+ compared to the GS3, and for the GS2 compared to the GS3 (CS+ > GS3, $p = .001$; GS2 > GS3, $p = .001$; all other ps > .09; CS+ vs. GS2, $p = .78$). In addition, for participants with higher IU, the LPP in acquisition training was descriptively smaller for the CS+ compared to the GS1, $p = .05$; all other ps > .06. No other effects reached significance at the omnibus level, all other ps > .14.

3.3.5 | LPP (2500–5000 ms)

No effects reached significance, all ps > .09.

4 | DISCUSSION

To shed light on opposing findings associated with IU, we conducted collaborative, direct and conceptual replications of the Morriss, Macdonald, et al. (2016) study, in which
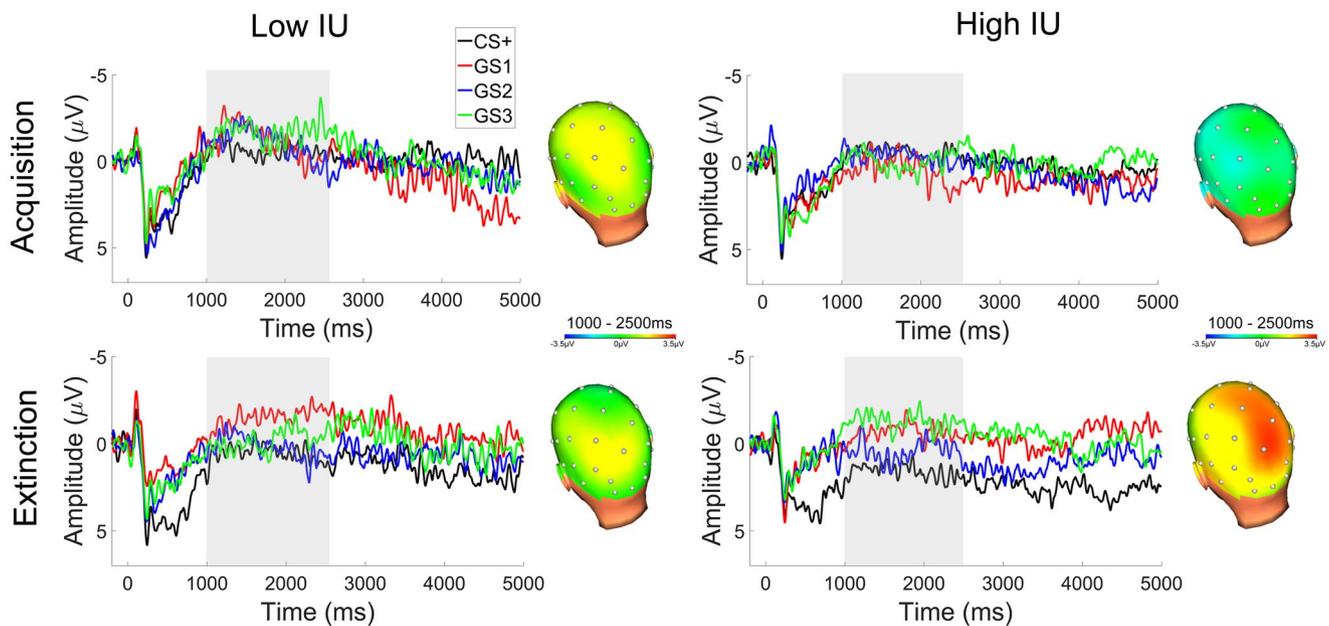

FIGURE 5 Grand-averaged waveforms (from conceptual replication 2) at the centroparietal pooling, where the LPP was scored, shown separately for each stimulus and phase, for participants with Low IU (1 *SD* below mean; left) and High IU (1 *SD* above mean; right). Headmaps depict the distribution of amplitude differences between CS+ and GS3 in each phase from 1,000 to 2,500 ms after stimulus onset. Omnibus analyses used continuous IU scores; "Low" and "High" IU groups were only created to follow-up and illustrate significant interactions. CS+ = conditioned stimulus; GS1, GS2, GS3 = generalization stimuli

higher IU participants demonstrated greater threat generalization during threat acquisition, as well as delayed learning of cue pairings. The original lab site attempted a direct replication, while two new lab sites attempted conceptual replications with extension to ERPs.

In the direct replication, we not only failed to replicate the interactions involving IU that Morriss, Macdonald, et al. (2016) reported, but we also observed additional findings that are hard to reconcile with the original study. For example, compared to lower IU participants, those with higher IU showed evidence of *less* threat generalization (i.e., greater differentiation between stimuli) across both training phases of the experiment (as measured using SCR). By contrast, Morriss, Macdonald, et al. (2016) had found that individuals with higher IU showed *greater* threat generalization (during acquisition training). Our inability to replicate the Morriss, Macdonald, et al. (2016) findings in the direct replication was unexpected, particularly given that this replication attempt was conducted in the same lab as the original study (but see Nelson et al., 2015). Yet it is worth considering that even given a true effect, a replication (like any study) has only limited power to detect if an effect exists—for example, with 80% power, 1/5 of replications will yield null results in the face of real effects (Anderson & Maxwell, 2017). Furthermore, no study is a true repetition of another, which necessarily creates uncertainty about failed replications. For example, was the first study a false positive? Or did the second study fail to replicate the conditions under which the effect was first observed/does it identify a boundary condition (Nosek & Errington, 2019)? Along these lines, different sample types might in part explain the discrepancy between our direct replication and Morriss, Macdonald, et al. (2016). That is, Morriss, Macdonald, et al. (2016) used an undergraduate sample, whereas the direct replication in the current study used a community sample with more diverse sample characteristics (e.g., a wider age range, a wider range of IU scores) which may have affected results.

Nonetheless, even when using an undergraduate sample in conceptual replication 2, results did not align consistently with the original study. For example, we found that compared to lower IU participants, higher IU participants showed increased threat generalization between stimuli across *both* training phases (SCR for conceptual replication 2), whereas Morriss, Macdonald, et al. (2016) found evidence of increased threat generalization in higher IU participants only in acquisition training, with an IU-related increase in threat *discrimination* evident in extinction. However, in considering these results, it is notable that "phase" provides a relatively crude distinction between early and late learning, which in reality takes place in a continuous fashion throughout acquisition and extinction training. Therefore, future work may wish to increase the number of trials to facilitate a

more fine-grained analysis of time, given that the stimulus meaning changes throughout acquisition and extinction training as learning takes place. Additionally, we note that ITIs for conceptual replication 2 (1,000–3,000 ms) were somewhat short for elicitation of SCR,[5] and could have affected SCR on the next trial. That is, if SCR was already elevated from the previous trial, SCR on the following trial could have been attenuated. However, given that trial order was randomized, this should not have modulated condition effects; moreover, there were few SCR nonresponders, suggesting that responses were not severely attenuated, at least broadly, across trials and participants.

Despite some differences in findings, overall, we found support at the conceptual level for increased threat generalization among higher IU individuals using SCR. As noted, higher IU was associated with greater threat generalization between stimuli across training phases (conceptual replication 2); in addition, higher IU was associated with less differentiation overall between acquisition and extinction training (conceptual replication 1). These findings indicate that the predictions that emerge from Morriss, Macdonald, et al. (2016) hold against new evidence that used a similar sample (unselected undergraduates) but a modified experimental design. In considering the mechanisms behind threat generalization in IU, either increased reactivity to the GS and/or blunted reactivity to the CS+ could be responsible for impaired threat discrimination (Figure 4). While several studies have suggested that anxiety and related constructs (e.g., IU) are associated with heightened reactivity to generalized stimuli (e.g., Morriss, Christakou, et al., 2016; Morriss, Macdonald, et al., 2016), blunted reactivity to threat cues has also been observed in high IU, possibly due to aberrant defensive responding or IU's inhibitory qualities (e.g., Nelson & Shankman, 2011). Unpacking the nature of these abnormalities may be important to improving treatment of disorders characterized by high IU. For example, if individuals high in IU show a reduced response to aversive stimuli, this could be problematic for exposure therapy, in which arousal may be a necessary precondition for extinction (Culver, Stoyanova, & Craske, 2012). On the other hand, exaggerated response to safe stimuli/GS might require another treatment approach entirely.

During extinction training, Morriss, Macdonald, et al. (2016) observed that individuals with higher IU showed *increased* differentiation between threatening and nonthreatening stimuli compared to lower IU individuals, suggesting delayed learning of cue pairings (i.e., extinction resembles acquisition). Here, we found some support for an association between higher IU and

---

[5]Some studies measuring SCR use paradigms with longer ITIs than we used here (e.g., 15,000 – 20,000 ms or 20,000 – 25,000 ms; Kredlow et al., 2017) while other studies use ITIs that are comparable to those used in conceptual replications 1 and 2 (e.g., 3,000 – 6,450 ms; Morriss, Christakou, & van Reekum, 2015).

increased threat differentiation using SCR in the direct replication, though this was observed across both acquisition and extinction training. Using the LPP (conceptual replication 2), we were able to replicate the Morriss, Macdonald, et al. (2016) finding of increased threat differentiation during extinction training specifically, suggesting delayed learning of cue pairings for higher IU individuals, as higher IU individuals did not differentiate between stimuli until extinction training. It seems possible that this effect might be attributable primarily to increased threat responding among high IU participants during early extinction training, in which extinction of cue pairings has yet to occur and acquisition learning is maximal.[6] Along these lines, future work may wish to include a longer acquisition training period to determine whether enhanced threat discrimination during extinction training might reflect *delayed* learning of threat cue pairings (i.e., with enough trials of acquisition, threat cue pairings would eventually be learned) or inability to extinguish threat (i.e., once an association between the CS+ and the US has been learned, it might be resistant to extinction). For instance, recent work has found that some anxiety phenotypes (e.g., generalized anxiety disorder) show poor threat discrimination across two sessions of threat acquisition training, whereas other phenotypes (e.g., panic disorder) show this only for the first session of acquisition training, suggesting the potential to overcome these deficits (Cooper, Grillon, & Lissek, 2018).

Interestingly, the sites that made improvements to the task design (e.g., modifications to stimulus presentation and ITIs) yielded results that were most similar to the original study (Morriss, Macdonald, et al., 2016). This raises the question as to what constitutes a replication. Traditionally, a replication repeats a study's procedure and observes whether the same results are found. However, even the most faithful replications will inherently differ from the original study in at least some respects, meaning that every replication is a test of generalizability. More importantly, the point of replication may not be to replicate a procedure, but to advance theory by putting *predictions* to the test using new evidence (Nosek & Errington, 2019; Zwaan, Etz, Lucas, & Donnellan, 2018). As such, a narrow, procedurally focused definition of replication may not be the optimal or only test of a theory, and might even cloak theories that are too vaguely defined to yield testable predictions beyond the original experiment (Nosek & Errington, 2019; Zwann et al., 2018). Given these and other considerations (e.g., reluctance on the part of researchers, journals and funders to invest in repetition), conceptual replications will likely play a key role in the next generation of the replication effort (Nosek & Errington, 2019).

In addition to testing the replicability of the Morriss, Macdonald, et al. (2016) prior findings on IU and threat generalization, the current study provides an opportunity to evaluate the unique contributions of ERPs, SCR, and subjective

ratings to understanding threat generalization. As in Morriss, Macdonald, et al. (2016), self-reported uneasiness and SCR were largest to the CS+ and decreased across the GS (see also Dunsmoor, Mitroff, & LaBar, 2009; Morriss, Christakou, et al., 2016; Nelson et al., 2015). By contrast, the LPP was enhanced to the CS+ with no differentiation between GS, in line with research from Nelson and colleagues (2015), which also examined these processes using the LPP. Therefore, SCR and self-reported uneasiness were sensitive to perceptual gradations of stimuli, whereas the LPP was not. Because SCR reflects sympathetic nervous system responding, it may be especially responsive to bottom-up, perceptually mediated processing of stimuli (Dawson, Schell, & Filion, 2017). On the other hand, long-lasting neural components, such as the LPP, may be more sensitive to subsequent, more controlled processing of stimuli, in which threat may be categorically differentiated from safety. Indeed, the high temporal resolution of ERPs may be useful in achieving a more precise understanding of the stages at which IU affects stimulus processing. For instance, associations with IU were only observed for the later portion of the LPP (i.e., 1,000–2,500 ms; conceptual replication 2), suggesting that IU may be primarily associated with alterations in top-down, elaborative stages of stimulus processing rather than initial salience detection. As such, simultaneous assessment of the measures collected here does not yield redundant information, but rather provides a multilevel perspective on threat processing across time and at different biological levels.

The internal consistency of SCR and the LPP was poor (i.e., $\alpha < .50$) for many conditions in the present study (Table 2), raising concerns about the suitability of these measures for the study of individual differences in the context of the current paradigm, at least when using an unselected sample (Hajcak, Meyer, & Kotov, 2017). This is especially problematic as threat acquisition and extinction tasks are commonly used in the psychophysiological literature to examine individual differences in anxiety and related pathology (e.g., trauma-related disorders). Like items on a questionnaire, more trials typically provide better internal consistency; here, trial counts were rather low, particularly compared to what is normally used to elicit visual stimulus-elicited ERPs. For example, in acquisition, there were only six even and six odd CS+ trials (for a total of twelve trials) and four even and four odd GS trials (for a total of eight trials).[7] Because acquisition and extinction training involve learning, changes in the meaning of stimuli throughout the course of the experiment (e.g.,

---

[6]The threat conditioning paradigm in the present study did not contain enough trials for early versus late extinction analyses for ERP components.

[7]The Spearman–Brown formula (Nunnally, 1978) provides a means of estimating how reliability would change if a test was lengthened. Using this formula, we can estimate what reliability might have been had there been more trials for each condition. For example, if there had been 30 CS + trials instead of 10 CS + trials during extinction training (in keeping with more typical trial counts in ERP research), the internal consistency for the 400–1,000 ms LPP elicited in this condition for conceptual replication 1 would have been 0.61 instead of 0.34.

as participants learn cue pairings) may also affect reliability. Future research may wish to address these concerns by including more trials to increase signal-to-noise ratio, but will need to weigh this against practical concerns such as task duration, particularly when optimizing tasks for slower responses, such as SCR, that require longer ITIs. Additionally, future research may wish to ensure that results are not driven by individual differences in the perceived salience of the US (e.g., by asking participants to rate the US).

Taken together, and although the direct replication failed, our results provide evidence at the conceptual level that compared to lower IU individuals, higher IU individuals may show greater threat generalization and delayed learning of threat cue pairings. Additionally, this multisite replication effort demonstrates how multiple psychophysiological measures can be used to provide complementary levels of information about threat processing as it relates to IU. As a collaborative team that included the authors of the original study, the current work benefitted from multiple perspectives and a nonadversarial approach. For example, our access to reviews of the original study meant that we were able to incorporate suggested design improvements into our attempted conceptual replications, while also maintaining the original paradigm in order to perform a more traditional, direct replication of the prior work. Going forward, a climate that motivates collaborative, rather than adversarial replication efforts will likely play a key role in the second wave of replication, as we work to move beyond "crisis" and toward a practice of "correcting."

ORCID

Elizabeth A. Bauer https://orcid.org/0000-0002-5521-7712
Annmarie MacNamara https://orcid.org/0000-0003-0399-973X
Anna Weinberg https://orcid.org/0000-0001-6110-096X
Jayne Morriss https://orcid.org/0000-0002-7928-9673

REFERENCES

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the "replication crisis": Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, *52*(3), 305–324. https://doi.org/10.1080/00273171.2017.1289361

Bradley, M. M., & Lang, P. J. (2007). *The international affective digitized sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual* (Technical Report B-3). Gainesville, Fl: University of Florida.

Carleton, R. N. (2016). Into the unknown: A review and synthesis of contemporary models involving uncertainty. *Journal of Anxiety Disorders*, *39*, 30–43. https://doi.org/10.1016/j.janxdis.2016.02.007

Cooper, S. E., Grillon, C., & Lissek, S. (2018). Impaired discriminative fear conditioning during later training trials differentiates generalized anxiety disorder, but not panic disorder, from healthy control participants. *Comprehensive Psychiatry*, *85*, 84–93. https://doi.org/10.1016/j.comppsych.2018.07.001

Culver, N. C., Stoyanova, M., & Craske, M. G. (2012). Emotional variability and sustained arousal during exposure. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(2), 787–793. https://doi.org/10.1016/j.jbtep.2011.10.009

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology*, *52*, 95–111. https://doi.org/10.1016/S0301-0511(99)00044-7

Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (4th ed., pp. 217–243). New York, NY: Cambridge University Press.

Dugas, M. J., Buhr, K., & Ladouceur, R. (2004). The role of intolerance of uncertainty in etiology and maintenance of generalized anxiety disorder. In R. G. Heimberg, C. L. Turk, & D. S. Mennin (Eds.), *Generalized anxiety disorder: Advances in research and practice* (pp. 143–163). New York, NY: Guilford Press.

Dunsmoor, J. E., Mitroff, S. R., & LaBar, K. S. (2009). Generalization of conditioned fear along a dimension of increasing fear intensity. *Learning & Memory*, *16*(7), 460–469. https://doi.org/10.1101/lm.1431609

Foti, D., Hajcak, H., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal-spatial PCA. *Psychophysiology*, *46*(3), 521–530. https://doi.org/10.1111/j.1469-8986.2009.00796.x

Gentes, E. L., & Ruscio, A. M. (2011). A meta-analysis of the relation of intolerance of uncertainty to symptoms of generalized anxiety disorder, major depressive disorder, and obsessive-compulsive disorder. *Clinical Psychology Review*, *31*(6), 923–933. https://doi.org/10.1016/j.cpr.2011.05.001

Greenberg, T., Carlson, J. M., Cha, J., Hajcak, G., & Mujica-Parodi, L. R. (2013). Neural reactivity tracks fear generalization gradients. *Biological Psychology*, *92*, 2–8. https://doi.org/10.1016/j.biopsycho.2011.12.007

Hajcak, G., Castille, C., Olvet, D. M., Dunning, J. P., Roohi, J., & Hatchwell, E. (2009). Genetic variation in brain-derived neurotrophic factor and human fear conditioning. *Genes, Brain & Behavior*, *8*, 80–85. https://doi.org/10.1111/j.1601-183X.2008.00447.x

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, *35*, 129–155. https://doi.org/10.1080/8756564090352650

Hajcak, G., Meyer, A., & Kotov, R. (2017). Psychometrics and the neuroscience of individual differences: Internal consistency limits between-subjects effects. *Journal of Abnormal Psychology*, *126*(6), 823–834. https://doi.org/10.1037/abn0000274

Jensen, D., Cohen, J. N., Mennin, D. S., Fresco, D. M., & Heimberg, R. G. (2016). Clarifying the unique associations among intolerance of

uncertainty, anxiety, and depression. *Cognitive Behaviour Therapy*, *45*(6), 431–444. https://doi.org/10.1080/16506073.2016.1197308

Kredlow, M. A., Pineles, S. L., Inslicht, S. S., Marin, M., Milad, M. R., Otto, M. W., & Orr, S. P. (2017). Assessment of skin conductance in African American and Non-African American participants in studies of conditioned fear. *Psychophysiology*, *54*(11), 1741–1754. https://doi.org/10.1111/psyp.12909

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy*, *28*, 487–494. https://doi.org/10.1016/0005-7967(90)90135-6

Miller, G. A., Gratton, G., & Yee, C. M. (1988). Generalized implementation of an eye movement correction procedure. *Psychophysiology*, *25*(2), https://doi.org/10.1111/j.1469-8986.1988.tb00999.x

Morriss, J., Christakou, A., & van Reekum, C. M. (2015). Intolerance of uncertainty predicts fear extinction in amygdala-ventromedial prefrontal cortical circuitry. *Biology of Mood & Anxiety Disorders*, *5*(1), 4. https://doi.org/10.1186/s13587-015-0019-8

Morriss, J., Christakou, A., & van Reekum, C. M. (2016). Nothing is safe: Intolerance of uncertainty is associated with compromised fear extinction learning. *Biological Psychology*, *121*, 187–193. https://doi.org/10.1016/j.biopsycho.2016.05.001

Morriss, J., Macdonald, B., & van Reekum, C. M. (2016). What is going on around here? Intolerance of uncertainty predicts threat generalization. *PLoS ONE*, *11*(5), e0154494. https://doi.org/10.1371/journal.pone.0154494

Nelson, B. D., & Shankman, S. A. (2011). Does intolerance of uncertainty predict anticipatory startle responses to uncertain threat? *International Journal of Psychophysiology*, *81*(2), 107–115. https://doi.org/10.1016/j.ijpsycho.2011.05.003

Nelson, B. D., Weinberg, A., Pawluk, J., Gawlowka, M., & Proudfit, G. H. (2015). An event related potential investigation of fear generalization and intolerance of uncertainty. *Behavior Therapy*, *46*(5), 661–670. https://doi.org/10.1016/j.beth.2014.09.010

Nosek, B. A., & Errington, T. M. (2019). What is replication? https://doi.org/10.31222/osf.io/u4g6t

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Schienle, A., Köchel, A., Ebner, F., Reishofer, G., & Schäfer, A. (2010). Neural correlates of intolerance of uncertainty. *Neuroscience Letters*, *479*(3), 272–276. https://doi.org/10.1016/j.neulet.2010.05.078

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P., & Jacobs, G. (1983). *Consulting Psychologists Press*. Palo Alto, CA.

Tanovic, E., Gee, D. G., & Joormann, J. (2018). Intolerance of uncertainty: Neural and psychophysiological correlates of the perception of uncertainty as threatening. *Clinical Psychology Review*, *60*, 87–99. https://doi.org/10.1016/j.cpr.2018.01.001

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. https://doi.org/10.1017/S0140525X17001972

SUPPORTING INFORMATION

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

How to cite this article: Bauer EA, MacNamara A, Sandre A, et al. Intolerance of uncertainty and threat generalization: A replication and extension. *Psychophysiology*. 2020;00:e13546. https://doi.org/10.1111/psyp.13546