Received: 16 August 2022 | Revised: 13 December 2022 | Accepted: 14 January 2023

DOI: 10.1111/psyp.14265

PSYCHOPHYSIOLOGY SPR WILEY

REVIEW

Multiverse analyses of fear acquisition and extinction retention in posttraumatic stress disorder

Michael W. Lewis[1,2] | **Daniel E. Bradford**[3] | **Edward F. Pace-Schott**[2,4] | **Scott L. Rauch**[1,2] | **Isabelle M. Rosso**[1,2]

[1]Center for Depression, Anxiety, and Stress Research, McLean Hospital, Belmont, Massachusetts, USA

[2]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA

[3]School of Psychological Sciences, Oregon State University, Corvallis, Oregon, USA

[4]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA

Correspondence
Michael W. Lewis and Isabelle M. Rosso, Center for Depression, Anxiety, and Stress Research, McLean Hospital, mailstop 334, Belmont, MA 02478, USA.
Email: mlewis@mclean.harvard.edu and irosso@hms.harvard.edu

Funding information
National Institute of Mental Health, Grant/Award Number: 5R01MH096987 (Rosso)

Abstract

Persistent fear is a cardinal feature of posttraumatic stress disorder (PTSD), and deficient fear extinction retention is a proposed illness mechanism and target of exposure-based therapy. However, evidence for deficient fear extinction in PTSD has been mixed using laboratory paradigms, which may relate to underidentified methodological variation across studies. We reviewed the literature to identify parameters that differ across studies of fear extinction retention in PTSD. We then performed Multiverse Analysis in a new sample, to quantify the impact of those methodological parameters on statistical findings. In 25 PTSD patients (15 female) and 36 trauma-exposed non-PTSD controls (TENC) (20 female), we recorded skin conductance response (SCR) during fear acquisition and extinction learning (day 1) and extinction recall (day 2). A first Multiverse Analysis examined the effects of methodological parameters identified by the literature review on comparisons of SCR-based fear extinction retention in PTSD versus TENC. A second Multiverse Analysis examined the effects of those methodological parameters on comparisons of SCR to a danger cue (CS+) versus safety cue (CS−) during fear acquisition. Both the literature review and the Multiverse Analysis yielded inconsistent findings for fear extinction retention in PTSD versus TENC, and most analyses found no statistically significant group difference. By contrast, significantly elevated SCR to CS+ versus CS− was consistently found across all analyses in the literature review and the Multiverse Analysis of new data. We discuss methodological parameters that may most contribute to inconsistent findings of fear extinction retention deficit in PTSD and implications for future clinical research.

KEYWORDS

fear extinction retention, fear learning, multiverse analysis, posttraumatic stress disorder, psychophysiology, skin conductance

1 | INTRODUCTION

Fear learning is among the most successful translational models in psychophysiology (Andero & Ressler, 2012). Extinction of conditioned fear is the basis of prolonged exposure therapy, one of the most effective psychotherapies for posttraumatic stress disorder (PTSD) (Rauch et al., 2012). However, about half of patients who undergo prolonged exposure do not recover (Ponniah & Hollon, 2009). Human laboratory studies can assess for deficient fear extinction retention using multi-day associative fear learning paradigms, most commonly using skin

 © 2023 Society for Psychophysiological Research.

conductance response (SCR) as an index of phasic electrodermal response to conditioned stimuli (Lonsdorf, Merz, et al., 2019). These paradigms have produced seminal contributions in the field due to their translational importance for PTSD (Milad et al., 2007, 2008, 2009) and for other anxiety disorders (e.g., Milad et al., 2013). However, analyses of fear extinction retention in PTSD have yielded inconsistent results across studies (Garfinkel et al., 2014; Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008, 2009; Orr et al., 2006; Pineles et al., 2016; Pöhlchen et al., 2020; Shvil et al., 2014). A precise and complete formulation of fear extinction retention as a translational model for exposure therapy in PTSD requires an understanding of potential moderating factors that may contribute to variability in laboratory findings (Fullana et al., 2020), including variability in methodological parameters. However, to our knowledge, variability in methods has not been specifically and systematically examined with respect to its influence on fear extinction retention findings in PTSD.

Using SCR as an index of fear, some studies of individuals with PTSD versus controls have found a statistically significant fear extinction retention deficit in PTSD (Garfinkel et al., 2014; Milad et al., 2008, 2009), whereas others have found no statistically significant group difference (Helpman et al., 2016; Marin et al., 2016; Orr et al., 2006; Pineles et al., 2016; Pöhlchen et al., 2020; Shvil et al., 2014). Between-study methodological differences have been raised as potential contributors to discrepant findings in studies of fear extinction retention PTSD. For example, Milad et al. (2008) posited that longer elapsed periods between fear extinction learning and fear extinction retention (e.g., 1 week; (Orr et al., 2006)) may make it less likely to identify retention deficits in PTSD than shorter periods between those phases (e.g., 24 h). Since then, however, multiple SCR studies found no fear extinction retention deficit 24 h after extinction learning in PTSD patients (Helpman et al., 2016; Marin et al., 2016; Pineles et al., 2016; Pöhlchen et al., 2020; Shvil et al., 2014). Thus, other methodological parameters may be driving discrepancies in the PTSD extinction retention field.

Evidence from the broader fear learning literature has started to identify methodological variations that affect outcomes of fear learning studies broadly that could also be impacting findings in clinical populations, including PTSD. For example, in seminal contributions to this area of inquiry, Lonsdorf and colleagues demonstrated that three influential methodological considerations in human fear conditioning paradigms are the exclusion of SCR non-responders, the exclusion of SCR non-learners, and variability in the formula used to operationalize SCR-based fear extinction retention (Lonsdorf et al., 2022; Lonsdorf, Klingelhöfer-Jens, et al., 2019; Lonsdorf, Merz, et al., 2019). With regard to the first factor, some but not all fear conditioning studies have excluded from their analyses participants who do not exhibit a detectable SCR to an aversive or threat-evoking stimulus (e.g., the unconditioned aversive stimulus or the conditioned danger cue); these participants are classified as "SCR non-responders (Lonsdorf, Klingelhöfer-Jens, et al., 2019)." With regard to the second factor, some but not all fear conditioning studies have excluded from their analyses participants who appear to not have learned the conditioning contingencies during fear acquisition; these participants are classified as "non-learners" because they do not develop a differential SCR response to the learned danger cue (CS+) versus the learned safety cue (CS−) during fear acquisition. With regard to the third factor, it has been common practice in the field to calculate a fear extinction retention index (ERI) from SCR data in order to quantify the amount of extinction learning maintained after a delay. An underlying assumption is that the ERI represents a standardized index that is comparable and replicable across studies. However, Lonsdorf and colleagues showed that there have been at least 16 different formulas for ERI in the SCR literature and that these variants show different correlations with each other when applied to a single data set, suggesting that they may not always reflect fear extinction retention in a similar way (and may be confounded by non-associative processes, see Lonsdorf, Merz, et al., 2019). In a recent study of PTSD and other fear-related disorders, Pöhlchen et al. (2020) identified removal of SCR outliers as another methodological parameter that may influence findings on fear extinction. Specifically, they found that a significant interaction of stimulus-by-group disappeared after the removal of participants with an average SCR response that deviated from the group average by more than 3.3 standard deviations. This outlier-dependent interaction suggested elevated SCR responding to the CS+ in a healthy control group compared with both a PTSD group and a transdiagnostic fear group that excluded PTSD, demonstrating possible relevance of this methodological factor to PTSD and other fear-related disorders (Pöhlchen et al., 2020). These studies collectively identify high methodological heterogeneity of fear learning and show that this heterogeneity can impact results and lead to different conclusions. They also motivate examination of the impact of methodological parameters on fear extinction retention findings in PTSD.

Multiverse Analysis is a novel approach to data analysis that can identify the extent to which methodological parameters create variance in results (Steegen et al., 2016). To perform a multiverse analysis, one first systematically identifies methodological variants that can be or have been used to test a particular hypothesis. Then, the multiverse analysis takes all combinations of the identified variants, producing a set of possible

analyses (or "universes") (Schweinsberg et al., 2021). For example, a multiverse analysis examining 16 variants of ERI, 3 variants of non-responder exclusion, and 2 variants of outlier removal would yield $16 \times 3 \times 2 = 96$ different universes. The output of a multiverse analysis quantifies and visually displays the results of all universes, thereby providing a representation of the degree to which findings vary across methodological permutations contained in the multiverse. Thus, multiverse allows interpretation of how robust the results of an analysis are to variability in a methodology or combination of methodologies (Steegen et al., 2016). Compared to traditional analytic approaches such as meta-analysis, multiverse analysis has the advantage of being able to determine the extent to which an effect is robust to different data-processing and analytic decisions within a single sample (LeBel et al., 2018; Steegen et al., 2016).

Multiverse analyses have been applied in recent studies of healthy samples to evaluate how results of within-subject analyses of fear acquisition and extinction are impacted by methodology. First, Kuhn et al. (2022) quantified SCR during fear acquisition using 8 different data processing pipelines.[1] In a follow-up study, Sjouwerman et al. (2022) performed a multiverse analysis of 605 different combinations of specific SCR quantification and data transformation decisions. During both fear acquisition and extinction, Lonsdorf et al. (2022) performed a multiverse analysis to examine potential impacts of the type of statistical model and the number of SCR trials used to quantify fear across 25 combinations of those two methodological decisions. These studies have evaluated the impact of methodology on results using three criteria: (1) consistency of detecting a "signal" across methodological combinations, defined as, for example $p <= .05$; (2) consistency of effect size estimates across methodological combinations; (3) precision of effect size estimates across methodological combinations (Kuhn et al., 2022; Lonsdorf et al., 2022; Sjouwerman et al., 2022). Collectively, these studies suggest high consistency of the finding that SCR is elevated to a CS+ compared with a CS− during fear acquisition in healthy adults, and that the consistency and precision of effect size estimates vary across different methodological combinations and samples. Compared to traditional analytic approaches that examine methodological factors individually, multiverse analysis has the advantage of also examining how combinations of methodological choices interact to impact findings (Schweinsberg et al., 2021).

One challenging aspect of interpreting a multiverse analysis is that it may be difficult to evaluate the impact of specific methods when examining across combinations of many methodological decisions (Liu et al., 2021). A common approach has been to descriptively identify patterns in multiverse findings to highlight impactful methodologies (e.g., Steegen et al., 2016). Although descriptive interpretations of multiverse results have led to important insights, it may be important to supplement this approach with an objective indicator of the impact of each specific methodological parameter (Liu et al., 2021). Thus, the recently developed Boba Sensitivity Analysis enables quantification of the specific impact of each methodology (Liu et al., 2021). However, this approach has yet to be applied to psychophysiological data (Schweinsberg et al., 2021). In the current report, we extend the application of Multiverse Analysis and Boba Sensitivity Analysis to the study of fear extinction retention in PTSD. In doing so, we aim to demonstrate the applicability of Multiverse Analysis and Boba Sensitivity Analysis to clinical group comparisons of fear learning, in addition to directly informing the PTSD literature.

The aims of this study were to conduct: (1) a literature search to identify consistencies and inconsistencies in methodology and results from past studies that compared SCR-derived fear extinction retention in PTSD versus trauma-exposed non-PTSD controls (TENC); (2) two Multiverse Analyses to determine the degree to which varying methodologies identified in the literature review impact fear acquisition and extinction retention results in a new data set. In the first Multiverse Analysis (*Acquisition Multiverse*), we examined fear acquisition. Based on evidence that conditioned fear acquisition is among the most reliable findings in psychophysiology (Vervliet & Boddez, 2020), we hypothesized (Hypothesis 1) that fear acquisition findings (CS+ > CS−) would be statistically significant across all combinations of methodological variants. In the second Multiverse Analysis (*Extinction Retention Multiverse*), we compared fear extinction retention in PTSD versus TENC. Based on the observation of inconsistent findings across prior studies of fear extinction retention in PTSD, we hypothesized (Hypothesis 2) that fear extinction retention findings would not be statistically significant across all combinations of methodological variants. Based on a lack of prior evidence for large and robust group differences in fear conditioning between individuals with fear-related disorders and controls (Pöhlchen et al., 2020), we had as a secondary hypothesis (Hypothesis 3) that *Extinction*

[1]Kuhn et al. (2022) labeled their analysis a "manyverse" and describe it as a "multiverse-type of approach." Their approach was similar to a multiverse analysis but was designed to explore a smaller set of methodologies in depth. Thus, they did not examine all possible combinations of SCR quantification methods identified by their literature search. For a detailed description of manyverse analysis and other "multiverse-style methods," see Kuhn et al. (2022) and others (e.g., Del Giudice & Gangestad, 2021; Lonsdorf et al., 2022; Sjouwerman et al., 2022).

Retention Multiverse findings would be less precise on average and less consistent across different methodological combinations when compared to *Acquisition Multiverse* findings. As Exploratory Analyses, we assessed the impact of each methodology using the recently-developed Boba Sensitivity Analysis (Liu et al., 2021).

2 | METHOD

2.1 | Literature search

We performed a systematic literature search to identify all SCR studies of fear extinction retention in PTSD published until December 2021. Within those studies, we examined all within-subjects analyses of fear acquisition (CS+ versus CS−) and all between-subjects analyses of fear extinction retention (PTSD versus TENC). We then identified methodologies for which researchers must make choices in sample composition or data processing methods, with each choice representing a possible variant of that methodology (Glossary, Table S1). Methodologies that met inclusion criteria were then classified as either consistent methodologies or inconsistent methodologies. Consistent methodologies were those for which a single variant was applied in all studies included in the literature review. Inconsistent methodologies were those for which different variants were applied across the studies in the literature review. Methodologies that could not be examined in multiverse analyses (Aim 2) on our existing data were excluded (e.g., we had a civilian sample so we could not include the decision to study a civilian versus military sample as a variant). Finally, we cataloged and synthesized results from the reviewed studies. For additional literature review details, see Supplement.

2.2 | New dataset

2.2.1 | Participants

Participants were 61 adults, including 25 participants with DSM-IV PTSD and 36 TENC. Table 1 shows participant demographic and clinical characteristics. PTSD diagnosis and Criterion A trauma exposure were assessed using the Clinician-Administered PTSD Scale for DSM-IV (CAPS; (Blake et al., 1995)) and the Life Events Checklist (LEC; (Gray et al., 2004)). TENC participants endorsed having experienced at least one DSM-IV PTSD criterion A trauma and had never met criteria for DSM-IV PTSD. The Institutional Review Board of McLean Hospital and the Partners Human Research Committee approved the study procedures. All participants provided written informed consent. For additional sample details, see Supplement.

TABLE 1 Demographic and clinical characteristics of the sample, mean ± standard deviation, or *N* (%)

Variable	PTSD ($n = 25$)	TENC ($n = 36$)	Group comparison
Age	35.43	33.14	$t = -1.11$, $p = .27$
Sex	15F, 10 M	20F, 16 M	$\chi^2 < 0.01$, $p = .93$
CAPS total	57.16 ± 17.25	7.69 ± 9.28	$t = 13.08$, $\mathbf{p < .01}$
LEC total	8.62 ± 3.35	6.74 ± 4.06	$t = 1.93$, $p = .06$
Ethnicity			$\chi^2 = 6.01$, $p = .20$
Asian	0 (0)	6 (17)	
Black	4 (16)	6 (17)	
Hispanic	1 (4)	0	
Other	2 (8)	2 (6)	
White	18 (72)	22 (61)	

Abbreviations: CAPS, Clinician Administered PTSD Scale for DSM-IV; F, female; LEC, Life Experiences Checklist; M, male; PTSD, posttraumatic stress disorder; TENC, trauma-exposed non-PTSD control. $p \leq .05$ shown in bold.

2.2.2 | Fear conditioning paradigm

We used a 2-day fear conditioning protocol as previously described (Helpman et al., 2016; Milad et al., 2009; Shvil et al., 2014). The unconditioned stimulus (US) was an electric shock to the finger. Conditioned stimuli (CSs) and context stimuli were pictures of different colored lamps and the rooms in which the lamps appeared, respectively. On day 1, participants completed *Habituation*, *Acquisition*, and *Extinction Learning*. During *Acquisition*, participants were exposed to 8 trials of subsequently extinguished CS+ (CS+E), 8 CS+ subsequently left unextinguished (CS+U), and 16 CS− trials; CS+E and CS+U were immediately followed by the US on 62.5% of trials. Approximately one minute later, during *Extinction Learning*, participants were exposed to 16 CS+E trials and 16 CS− trials, neither of which were followed by the US and both of which were presented in a different context than the acquisition context. On day 2, participants completed *Extinction Recall*, which consisted of 8 CS+E trials, 8 CS+U trials, and 16 CS− trials, presented in the *Extinction Learning* context, none of which were followed by the US. During *Acquisition* and *Extinction Recall*, CS+E and CS+U were presented in blocks, with each block containing 8 trials of either CS+ type (8 CS+E or 8 CS+U) intermixed with 8 CS− trials, for a total of 16 trials per block, counterbalanced between subjects. For schematic, see Figure 1. For additional paradigm and data collection details, see Supplement.

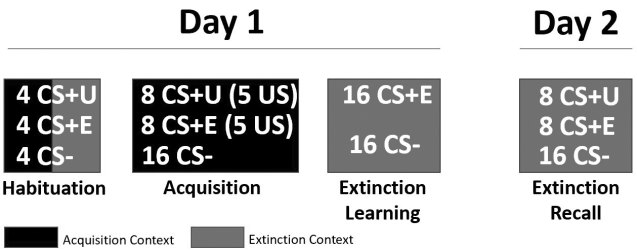

FIGURE 1 Schematic of fear learning task. Adapted from Milad et al. (2005). CS+U = danger cue left unextinguished, CS + E = danger cue extinguished, CS− = safety cue, US = unconditioned stimulus.

2.2.3 | Data collection

Skin conductance level (SCL) was detected from two 9-mm Ag/AgCl radiotranslucent electrodes (BioPac Systems Inc., Goleta, CA) that were filled with isotonic paste, separated by 14 mm, and placed on the hypothenar surface of the participant's nondominant hand (contralateral to the hand receiving the shock) as recommended by published guidelines (Fowles et al., 1981). SCL was directly recorded at a rate of 10 Hz by a Coulbourn Isolated Skin Conductance coupler (S71-23, Coulbourn Instruments, Allentown, PA) that applied a constant voltage of 0.5 V and was expressed in microSiemens. No filtering was applied. The analog signals were digitized using a Coulbourn Lablinc Analog-to-Digital Converter (V19–16). Skin conductance response (SCR) to all conditioned and unconditioned stimuli was calculated using Coulburn software by subtracting mean SCL during the final 2 s of context alone from the highest SCL during the 6-s CS presentation, as described previously (Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008, 2009; Shvil et al., 2014). No minimum response threshold was applied. All SCR trials that were not lost due to recording error were included in analyses. SCR trials lost due to recording error (e.g., electrode detachment, excessive baseline activity, responses outside the sampling window) were treated as missing and unreplaced. Across all participants and trials, 8.45% of trials were lost due to recording error. This percentage is in line with the only study included in our review that reported the percentage of missing trials (Pineles et al., 2016). Two participants were excluded due to technical errors during the experiment.

2.2.4 | Multiverse analysis

We performed one Multiverse Analysis for each of our primary hypotheses: (1) fear acquisition findings (CS+ > CS−) will be consistently statistically significant across all combinations of methodological variants; (2) fear extinction retention findings (PTSD < TENC) will not be consistently statistically significant across all combinations of methodological variants. Each multiverse analysis entailed performing t-tests across all combinations of inconsistent methodologies (Tables 2 and 3); each combination represented one "Universe." Consistent methodologies were applied uniformly to all Universes. Cohen's d was calculated for each t-test. To

TABLE 2 Inconsistent methodologies to test fear acquisition (CS+ versus CS−)

Inconsistent Methodology 1: Acquisition Quantification (AQ)
AQ1: [mean (First 4 CS + E *Acquisition*, First 4 CS + U *Acquisition*)] vs. [mean (First 4 CS− trials that occurred during the same block as the CS + E *Acquisition*, First 4 C- trials that occurred during the same block as the CS + U *Acquisition*)]
AQ2: [mean (All 8 CS + E *Acquisition*, All 8 CS + U *Acquisition*)] vs. [mean (All 16 CS− *Acquisition*)]
AQ3: [mean (Trials 2–5 CS+ *Acquisition*)] vs. [mean (Trials 2–5 CS− *Acquisition*)]
AQ4: [mean (Unreinforced CS+ Trials *Acquisition*)] vs. [mean (Paired CS− Trials *Acquisition*)]
AQ5: [mean (Trials 4–5 CS+ *Acquisition*)] vs. [(mean (Trials 4–5 CS− *Acquisition*)]
Inconsistent Methodology 2: Sample Sex
F + M: Study sample includes both female and male participants
F: Study sample includes female participants only
M: Study sample includes male participants only
Inconsistent Methodology 3: SCR Non-responder Exclusion
<0.05 uS: Excluded if SCR to US <0.05 uS
<0.1 uS: Excluded if SCR to US <0.1 uS
NR/No: Not Reported/No SCR Non-responder Exclusion[N]
Inconsistent Methodology 4: CAPS Exclusion
Yes: Exclude PTSD participants if CAPS <50, Exclude TENC participants if CAPS >19
NR/No: Not Reported / No CAPS Exclusion[N]
Inconsistent Methodology 5: Outlier Trial Removal
> 3 *SD*: > 3 Standard Deviations From Mean
NR/No: No Outlier Trial Removal[N]
'screen': "Screened for Outliers"[S]

Abbreviations: CAPS, Clinician Administered PTSD Scale for DSM-IV; CS+, learned danger cue; CS−, learned safety cue; CS + E, learned danger cue subsequently extinguished; CS + U, learned danger cue left unextinguished; F, Female; M, Male; [N], Not Reported in cited study, Multiverse Analysis includes a variant of no exclusion / removal; NR, not reported; [S], Not included in Multiverse Analysis due to insufficient information; PTSD, posttraumatic stress disorder; TENC, trauma-exposed non-PTSD control; US, unconditioned stimulus.

facilitate interpretation, all t-tests were coded so that a positive Cohen's d indicated a mean difference in the predicted direction of the hypothesis and prevailing theory (e.g., CS+ > CS− for the *Acquisition Multiverse*); conversely, a negative Cohen's d indicated a mean difference in the opposite direction of the hypothesized finding and prevailing theory (e.g., greater fear extinction retention in PTSD than TENC for the *Extinction Retention Multiverse*). We evaluated the consistency and precision of results for each Multiverse Analysis by applying a recently developed framework, described in more detail in the "Framework" subsection (Kuhn et al., 2022; LeBel et al., 2018; Lonsdorf et al., 2022; Sjouwerman et al., 2022).

We used Boba DSL (Liu et al., 2021), JSON (Pezoa et al., 2016), and R (R Core Team, 2013) to perform multiverse analysis. We wrote the multiverse analysis code in R Studio and used Boba DSL to generate R code for each variant. We wrote a JSON file to compile all results. Boba commands were executed in the R software environment using Conda (Boroumand et al., 2019) in the command prompt (see Liu et al., 2021 for tutorial). Multiverse analysis figures were created in R using ggplot version 3.2.1 (Wickham, 2016). All t-tests in the *Conditioning Multiverse* used SCR as the dependent variable and stimulus (CS+ versus CS−) as the independent variable; CS+ and CS− were quantified using one of the five variants of Acquisition Quantification (Table 3). All analyses in the *Extinction Retention Multiverse* used one of the seven variants of ERI (Table 4) as the dependent variable and diagnostic group (PTSD versus TENC) as the independent variable. All Cohen's d calculations were performed using Hedges' correction to reduce bias introduced by small samples (Grissom & Kim, 2005).

We assessed the impact of each methodology using the recently-developed Boba Sensitivity Analysis

TABLE 3 Inconsistent methodologies to compare fear extinction retention (PTSD versus TENC)

Inconsistent Methodology 1: Extinction Retention Index (ERI)
ERI1: 100 − [100 × mean (first 4 CS + E *Extinction Retention*)/max (CS + E *Acquisition*)]
ERI2: 100 − [100 × mean (first 2 CS+ *Extinction Retention*)/max (CS+ *Acquisition*)]
ERI3: [mean (first 5 CS+ *Extinction Retention*) − mean (first 5 CS− *Extinction Retention*)] − [mean (trial 2–5 CS+ *Extinction Learning*) − mean (trial 2–5 CS− *Extinction Learning*)]
ERI4: [mean (first 5 CS+ *Extinction Retention*) − mean (first 5 CS− *Extinction Retention*)] − [mean (last 5 CS+ *Extinction Learning*) − mean (last 5 CS− *Extinction Learning*)]
ERI5: [mean (first 4 CS + E *Extinction Retention*) − (mean (first 4 CS + U *Extinction Retention*)]
ERI6: [mean (first 4 CS + E *Extinction Retention*) − (mean (first 4 CS− *Extinction Retention*)]
ERI7: [mean (first 2 CS + E *Extinction Retention*) − (mean (first 2 CS− *Extinction Retention*)]
Inconsistent Methodology 2–5: See Table 2

Abbreviations: CS+, learned danger cue; CS−, learned safety cue; CS + E, learned danger cue subsequently extinguished; CS + U, learned danger cue left unextinguished; PTSD, posttraumatic stress disorder; TENC, trauma-exposed non-PTSD control.

TABLE 4 Results of analyses of fear acquisition (CS+ > CS−)

Study information		**Inconsistent methodologies**					**Results**	
Citation	**Sample size**	**Acquisition quantification**	**Sex**	**Non-responder**	**CAPS**	**Outlier**	**Effect size**	***p*-value**
Milad et al. (2009)	$n = 31$	AQ1	M + F	NR	NR	NR	$F = 19.6$ [NR]	$\mathbf{p < .001}^{***}$
Shvil et al. (2014)	$n = 56$	AQ1	M + F	NR	Yes	NR	$F = 27.64$ [NR]	$\mathbf{p < .001}^{***}$
Helpman et al. (2016)	$n = 47$	AQ2	M + F	<0.05 uS	Yes	NR	$F = 57.48$ [NR]	$\mathbf{p < .001}^{***}$
Milad et al. (2008)	$n = 28$	AQ3	M	NR	NR	NR	$F = 8.6$ [NR]	$\mathbf{p = .014}^{p*}$
Garfinkel et al. (2014)	$n = 21$	AQ4	M	NR	NR	>3 *SD*	$F = 8.88$ [NR]	$\mathbf{p = .007}^{**}$
Pineles et al. (2016)	$n = 32$	AQ5	F	<0.1 uS	NR	'screen'^{s}	$B = 0.40$ [0.09,0.50]	$\mathbf{p = .004}^{**}$

Note: All *p* values two-tailed. $p \leq .05$ shown in bold.

Abbreviations: AQ, Acquisition Quantification; CAPS, Clinician Administered PTSD Scale for DSM-IV; F, Female; M, Male; NR, Not Reported; ^{p}, *p* value reported in cited manuscript was one-tailed; ^{s}, Cited study "screened for outliers," no additional information specified.

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

(Liu et al., 2021). This applies one-way ANOVA to quantify the degree to which average Universe effect size estimates vary across variants of an Inconsistent Methodology. The F statistic represents a Sensitivity Score (Glossary, Table S1) and is interpreted descriptively because statistical assumptions needed to calculate a *p*-value are not met (Liu et al., 2021). To further evaluate the impact of specific Inconsistent Methodologies, we examined the stratified distributions of Universes meeting a $p \leq .05$ threshold, Cohen's d point estimates, and corresponding 95% CIs.

The Sensitivity Score for each Inconsistent Methodology was calculated using one-way ANOVA. In this case, the ANOVA F statistic provides an effect size for the impact of each Inconsistent Methodology on the average Cohen's d within a Multiverse Analysis. With this approach, each Universe acts as a participant/observation. An ANOVA was conducted with the Cohen's d as the dependent variable, with the Inconsistent Methodology as the factor, and the variants of the Inconsistent Methodology as the levels of the factor. The F value from the Sensitivity Analysis can be used to evaluate the effects of each Inconsistent Methodology descriptively, but since the Universes are not independently sampled, statistical assumptions needed for comparison to a critical F and generation of a *p*-value are not met (Liu et al., 2020).

2.2.5 | Framework

First, to evaluate the *consistency of detecting a "signal,"* we calculated the percentage of Universe results meeting a threshold of $p \leq .05$. Second, to evaluate the *precision* of effect size estimates across methodological combinations, we calculated the average 95% confidence intervals. Third, to evaluate the *consistency of effect size estimates* across methodological combinations, we calculated the percentage of results that met previously defined criteria for consistency (Kuhn et al., 2022; LeBel et al., 2018; Sjouwerman et al., 2022). Briefly, to determine whether two Universes were consistent, we examined whether the 95% confidence interval of any one Universe included the point estimate of the other. Fourth, as an additional measure of *consistency of effect size estimates*, we examined the distribution of Cohen's d point estimates. We included this measure because previous fear learning multiverse studies (Kuhn et al., 2022; Lonsdorf et al., 2022; Sjouwerman et al., 2022) found that multiverse analyses with less precise effect sizes estimates (i.e., larger CIs) tended to have higher consistency as measured by calculating the percentage of effect sizes falling within the CIs and noted this as a limitation of the measure or a potential source of bias (i.e., it may cause overestimation of consistency if CIs are very large). Thus, examination of the distribution of effect sizes (e.g., the effect size range as measured by the difference between the highest and lowest Cohen's d, whether the effect sizes are consistently in the predicted direction, size of the average effect relative to the range of effects) is an important aspect of evaluating the consistency of the effect size estimates. Also, evaluation of the distribution is in line with recommendations from other multiverse papers (Liu et al., 2021; Patel et al., 2015; Schweinsberg et al., 2021). Specifically, we examined the effect size *range* (as measured by the difference between the highest and lowest Cohen's d), effect size *standard deviation*, size of the average effect relative to the range of effects, and whether the effect sizes are consistently in the predicted *direction* (i.e., percentage positive). Regarding the effect size range and standard deviation, larger scores indicate less consistency. Regarding the average effect relative to the range, if the average effect is near zero but the range is large, this indicates inconsistency. With regard to effect size direction (e.g., positive or negative Cohen's d), the observation that the effect can go in either direction has been established as a strong signal of inconsistency across methodologies (Liu et al., 2021; Patel et al., 2015; Schweinsberg et al., 2021).

2.3 | Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures, and we follow Journal Article Reporting Standards (JARS). All analysis code and data are available at (https://osf.io/rybae/). The fear conditioning paradigm files are not on the repository because they are property of their original owners (Milad et al., 2009). Data were analyzed using R, version 4.0.0 (R Core Team, 2020) and Boba DSL (Liu et al., 2020). This study's design and analyses were not pre-registered.

3 | RESULTS

3.1 | Literature review

Our review identified 7 studies (Garfinkel et al., 2014; Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008, 2009; Pineles et al., 2016; Shvil et al., 2014), including 6 that tested whether differential fear acquisition occurred during *Acquisition* (Garfinkel et al., 2014; Helpman et al., 2016; Milad et al., 2008, 2009; Pineles et al., 2016; Shvil et al., 2014). Across all studies, we identified 2 Consistent Methodologies: (1) all studies excluded participants who met criteria for only Lifetime PTSD (Lifetime PTSD Exclusion); (2) all studies used square root transformation

to normalize SCR trials (SCR Transformation). We identified 5 inconsistent methodologies for *Acquisition* (Table 2) and 5 inconsistent methodologies for *Extinction Retention* (Table 3).

All 6 *Acquisition* analyses (100%) found significantly higher SCR responding to CS+ trials versus CS− trials (Table 4). Of 12 *Extinction Retention* analyses, 9 found no significant difference in fear extinction retention in PTSD versus TENC (75%); 6 found non-significantly lower fear extinction retention in PTSD (50%), 2 found non-significantly higher fear extinction retention in PTSD (16.7%), 1 reported only that the difference was not significant (8.3%), 3 found significantly lower fear extinction retention in PTSD (25%) (Table 5).

3.2 | Multiverse analysis of new data

3.2.1 | Multiverse structure

In the *Acquisition Multiverse*, we applied the 5 Inconsistent Methodologies listed in Table 2,[2] creating 180 Universes: (5 variants of Acquisition Quantification) × (3 variants of Sample Sex) × (3 variants of SCR Non-Responder Exclusion) × (2 variants of CAPS Exclusion) × (2 variants of Outlier Trial Removal).

In the *Extinction Retention Multiverse*, we applied the 5 Inconsistent Methodologies listed in Table 3,[2] creating 252 universes: (7 variants of ERI) × (3 variants of Sample Sex) × (3 variants of SCR Non-Responder Exclusion) × (2 variants of CAPS Exclusion) × (2 variants of Outlier Trial Removal).

3.2.2 | Multiverse results

3.2.2.1 | *Acquisition multiverse*

All 180 Universes (100%) found significantly higher SCR responding to CS+ trials versus CS− trials during acquisition (mean p-value = .0003). Across all pairwise comparisons between Universes in the *Acquisition Multiverse*, the Cohen's d point-estimate of one Universe was within the 95% CI of the comparison Universe 94.3% of the time (30,371 out of 32,220). Figure 2a displays the distribution of Cohen's d point estimates and 95% confidence intervals within the *Acquisition Multiverse*. For variants and results of each Universe, see (https://osf.io/rybae/).

Based on the Sensitivity Score, CAPS Exclusion had the greatest impact on the Cohen's d point estimate ($F = 46.41$), followed by Sample Sex ($F = 10.37$), Acquisition Quantification ($F = 10.01$), SCR Non-responder Exclusion ($F = 9.89$), and Outlier Trial Removal ($F = 3.73$). For findings within the *Acquisition Multiverse*, stratified by Inconsistent Methodology, see Supplement.

3.2.2.2 | *Extinction retention multiverse*

Of 252 Universes, 214 found no significant difference in fear extinction retention in PTSD versus TENC (84.9%); 47 found *non-significantly lower* fear extinction retention in PTSD (18.6%), 167 found *non-significantly higher* fear extinction retention in PTSD (66.3%), 0 found *significantly lower* fear extinction retention in PTSD (0%), 38 found *significantly higher* fear extinction retention in PTSD (15.1%) (mean p-value = .41).[3] Across all pairwise comparisons between Universes in the *Extinction Retention Multiverse*, the Cohen's d point-estimate of one Universe was within the 95% confidence interval of the comparison Universe 88.4% of the time (55,885 out of 63,252). Figure 2b displays the distribution of Cohen's d point estimates, 95% confidence intervals, and findings of $p \leq .05$ within the *Extinction Retention Multiverse*. For variants and results of each Universe, see (https://osf.io/rybae/).

Based on the Sensitivity Score, Sample Sex had the greatest impact on the Cohen's d point estimate ($F = 49.57$), followed by ERI ($F = 32.81$), Outlier Trial Removal ($F = 13.45$), CAPS Exclusion ($F = 3.88$), and SCR Non-responder Exclusion ($F = 0.34$). For findings within the *Extinction Retention Multiverse*, stratified by Inconsistent Methodology, see Supplement.

4 | DISCUSSION

Previous research has identified deficient fear extinction retention as a feature of PTSD and anxiety-related disorders, which may reflect a role in pathogenesis and/or treatment response to extinction-based cognitive behavioral therapy (CBT). Moreover, it has been proposed that extinction retention deficits in PTSD and anxiety

[2]Of the 7 studies in the literature review, 5 did not report whether they applied SCR Non-responder Exclusion (71%) (Garfinkel et al., 2014; Marin et al., 2016; Milad et al., 2008, 2009), 5 did not report whether they applied CAPS Exclusion (71%) (Garfinkel et al., 2014; Marin et al., 2016; Milad et al., 2008, 2009; Pineles et al., 2016), and 5 did not report whether they applied Outlier Trial Removal (71%) (Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008; Shvil et al., 2014). To account for those studies, we included the following variants in both Multiverse Analyses: No SCR Non-responder Exclusion, No CAPS Exclusion, and No Outlier Trial Removal (Tables 2 and 3). One study reported that they "screened for outliers" (Pineles et al., 2016); in the absence of further details, we could not include this variant in Multiverse Analysis.

[3]To address concerns about statistical power, we also performed supplemental analysis conducting each Universe with a one-tailed t-test. This doubles statistical power but did not change the pattern of results (see Supplement).

TABLE 5 Results of group comparisons of fear extinction retention (PTSD versus TENC)

Study information			Inconsistent methodologies					Results	
	Sample Size by Group								
Citation	PTSD	TENC	Extinction Retention Index	Sex	Non-Responder	CAPS	Outlier	Effect Size	*p*-value
Milad et al. (2009)	$n = 16$	$n = 15$	ERI1	M + F	NR	NR	NR	$t = 2.9$ [NR]	**$p < .02$***
Shvil et al. (2014)	$n = 31$	$n = 25$	ERI1	M + F	NR	Yes	NR	$t = 0.9$ [NR]	$p = .14^{p}$
Marin et al. (2016)	$n = 24$	$n = 20$	ERI1	M + F	NR	NR	NR	$t = 1.79$ [NR]	$p = .09$
Milad et al. (2008)[1]	$n = 7$	$n = 7$	ERI2	M	NR	NR	NR	$F = 3.0$ [NR]	$p = .12^{p}$
Pineles et al. (2016)	$n = 16$	$n = 16$	ERI3	F	< 0.1 uS	NR	'screen's	$B = -.15$ [−0.37,0.08]	$p = .20$
Pineles et al. (2016)	$n = 16$	$n = 16$	ERI4	F	< 0.1 uS	NR	'screen's	$B = -.04$ [−0.27,0.19]	$p = .74$
Helpman et al. (2016)	$n = 24$	$n = 23$	ERI5	M + F	< 0.05 uS	Yes	NR	NR [NR]	$p = .13$
Shvil et al. (2014)	$n = 31$	$n = 25$	ERI5	M + F	NR	Yes	NR	$F = 1.18$ [NR]	$p = .28$
Milad et al. (2009)	$n = 16$	$n = 15$	ERI5	M + F	NR	NR	NR	$F = 4.99$ [NR]	**$p = .03$***
Shvil et al. (2014)	$n = 31$	$n = 25$	ERI5	M + F	NR	Yes	NR	$F = 1.25$ [NR]	$p = .12$
Garfinkel et al. (2014)	$n = 12$	$n = 11$	ERI6	M	NR	NR	>3 *SD*	$F = 6.38$ [NR]	**$p = .02$***
Milad et al. (2008)[1]	$n = 7$	$n = 7$	ERI7	M	NR	NR	NR	NR [NR]	$p > .10^{p}$

Note: All *p* values two-tailed. p ≤ .05 shown in bold.

Abbreviations: ERI, Extinction Retention Index; F, Female; M, Male; NR, Not Reported; [p], *p* value reported in cited manuscript was one-tailed; PTSD, posttraumatic stress disorder; [s], Cited study reported "screened for outliers," no additional information provided; *SD*, Standard Deviation; TENC, trauma-exposed non-PTSD controls.

*$p < .05$.

[1]Total study sample was 28 (7 PTSD, 7 TENC, 14 nontraumatized healthy controls).

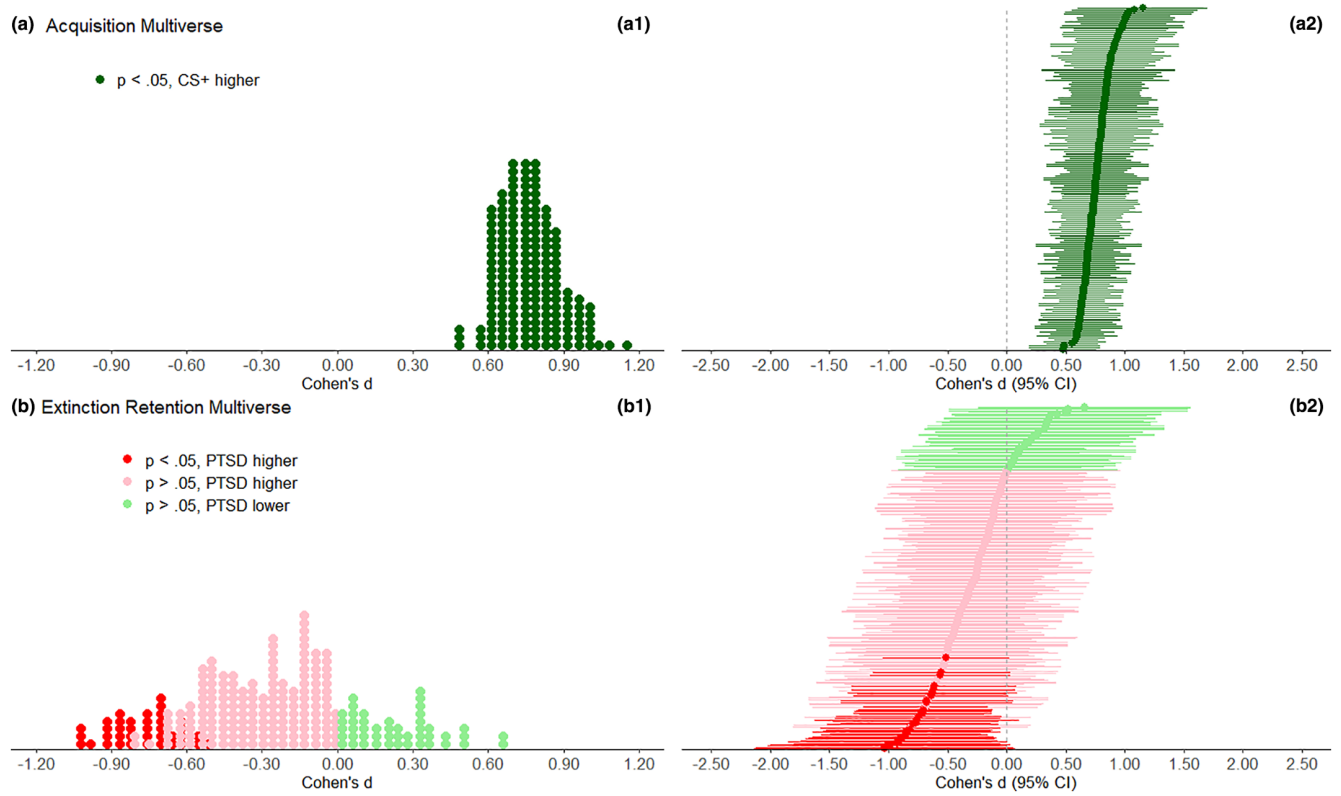

FIGURE 2 Distribution of effect sizes in the multiverse analyses. (a) Acquisition Multiverse; (A1: mean Cohen's d = 0.79, *SD* = 0.12) and (b) Extinction Retention Multiverse (B1: mean Cohen's d = −0.29, *SD* = 0.36). Also pictured are Confidence Interval (CI) Widths of the Acquisition Multiverse in A2 (mean 95% CI width = 0.73) and Extinction Retention Multiverse in B2 (mean 95% CI width = 1.64). Each dot indicates one Universe. PTSD = posttraumatic stress disorder, *SD* = standard deviation.

disorders could be a target for therapeutic intervention, whereby enhancing extinction retention capabilities could facilitate enhanced CBT response. However, laboratory findings have been inconsistent. This is the first study to identify methodological parameters that contribute to variability in fear extinction retention findings when comparing PTSD patients to trauma-exposed controls. In our Extinction Retention Multiverse, the use of different methodological combinations identified by a review of prior studies produced inconsistent findings within a single novel data set. Effect sizes ranged from a very large effect indicating higher average fear extinction retention in PTSD versus TENC (Cohen's d = −1.08) to a medium effect indicating lower fear extinction retention in PTSD versus TENC (d = 0.68). Overall, these findings demonstrate that methodological decisions that have varied across previous studies can lead to different results regarding fear extinction retention deficits in PTSD. Although the current study is not intended to refute the hypothesis of deficient fear extinction retention in PTSD, these findings demonstrate that an increased understanding of the impact of methodological parameters on statistical findings may be needed to refine translational models and to develop a more complete theory of fear extinction retention in PTSD that accounts for moderating factors. Importantly, this concern also applies to other clinical fear learning hypotheses (Fullana et al., 2020; LeBel et al., 2018; Lonsdorf & Merz, 2017; Tackett et al., 2019), and the current study establishes multiverse analysis as a promising tool for evaluating the credibility of clinical fear learning hypotheses.

Our findings underscore the potential for underreported methodological decisions to impact results in studies of fear extinction retention in PTSD. Our review found that most studies of fear extinction retention in PTSD did not report their methodology for Outlier Trial Removal (Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008, 2009; Pineles et al., 2016; Shvil et al., 2014). However, in the *Extinction Retention Multiverse*, analyses that excluded outliers were 2.8 times more likely to find a statistically significant difference between PTSD and TENC groups. Similarly, our review found that most studies did not report methodology on Non-Responder Exclusion (Garfinkel et al., 2014; Marin et al., 2016; Milad et al., 2008, 2009; Shvil et al., 2014) or CAPS Exclusion (Garfinkel et al., 2014; Marin et al., 2016; Milad et al., 2008, 2009; Pineles et al., 2016). Yet, each of these methodologies had a non-zero impact on effect size estimates and statistical significance in the *Extinction*

Retention Multiverse. The implications of these methodological decisions and their lack of reporting transparency extend beyond study replicability and implications for PTSD treatment to broad and serious challenges for outcome generalizability and science equity. For example, researchers disproportionally exclude Black participants as "non-responders" as Black participants often have lower SCR compared with White participants (Bradford et al., 2022; Kredlow et al., 2017; Webb et al., 2022). This worsens inexcusably wide racial disparities in participant recruitment and retention, and potentially interacts with other methodological decisions and demographics to decrease the representativeness of psychophysiological data (Bradford et al., 2022). In sum, our findings demonstrate the need for consensus reporting in the study of fear extinction retention in PTSD and are in line with previous calls for more consistent methodological reporting in psychological science (Bradford et al., 2022; LeBel et al., 2018; Webb et al., 2022).

The importance of Outlier Trial Removal for results of the *Extinction Retention Multiverse* extends previous findings in other populations (Morís Fernández & Vadillo, 2020; Pöhlchen et al., 2020). Pöhlchen et al. (2020) found that an initially significant group difference (PTSD versus other fear-related disorders versus healthy controls) that was driven by elevated SCR among controls during extinction learning was no longer significant after removing participants who were outliers. Similarly, in a simulated data set, Morís Fernández and Vadillo (2020) demonstrated that flexibility in methods for removing outlier trials can lead to a false-positive rate of 17% when alpha is .05, suggesting that this finding likely has relevance for clinical comparisons beyond PTSD and other fear-related disorders. Overall, this evidence suggests that it is important for future SCR studies of fear extinction retention in clinical populations such as PTSD to test for robustness to outliers.

Our finding of an impact of Extinction Retention Index (ERI) extends a landmark study by Lonsdorf, Merz, and colleagues (2019) in two ways: (1) we show that this methodological parameter interacts with other methodological choices to impact extinction retention findings; (2) we demonstrate that the operationalization of fear extinction retention is important for understanding PTSD, specifically. In our review, we focused specifically on studies of fear extinction retention in PTSD and identified 7 different operationalizations of fear extinction retention across 7 PTSD studies. Previously, Lonsdorf, Merz, et al. (2019) had found 16 different ways of quantifying fear extinction retention in their literature review comprising 37 studies. They had found inconsistent correlations of the 16 identified measures of extinction retention with each other and with total score on the Fear Survey Questionnaire (FSQ) using 4 different data sets (3 healthy samples and 1 subclinical phobia sample). In our *Extinction Retention Multiverse*, we found that variants of ERI had a large impact on findings of group comparisons of fear extinction retention (PTSD versus TENC). This extends prior work by demonstrating that ERI is important to consider for PTSD specifically and raises the possibility of its relevance to clinical group comparisons more broadly. Further, ERI interacted with other methodological decisions derived from our literature review that are relevant beyond the PTSD literature. For example, in the *Extinction Retention Multiverse*, the second ERI variant (ERI2) only led to a significant finding when two specific conditions were met: (1) the sample was all female (Sex = F), and (2) no symptom exclusion was applied (CAPS = No). Conversely, the sixth ERI variant (ERI6) only led to a significant finding when both females and males were included in the sample (Sex = M + F), and a symptom severity exclusion was applied (CAPS = Yes). More broadly, Figure S7 shows that the shape of the distribution of effects varies across the seven ERI variants, suggesting interactions with other methodological parameters (Liu et al., 2021; Schweinsberg et al., 2021). Overall, these findings show that ERI is an important methodological consideration for PTSD studies and that it interacts with other methodological parameters that are relevant beyond the PTSD literature.

Close examination of our ERI findings suggests a need for future studies to further pull apart noise from potentially etiologically relevant processes. As previously reviewed by Lonsdorf, Merz, et al. (2019), the large impact of Fear Extinction Retention Quantification variants on results may reflect differences in what learning mechanisms are measured (see Lonsdorf, Merz, et al., 2019 for detailed review). For example, of the 7 ERI variants in our review, only ERI3 and ERI4 normalized based on extinction learning (i.e., both subtracted an extinction learning difference score from an extinction retention difference score). Thus, it is surprising that the distributions of findings from Universes that used ERI3 and ERI4 had relatively little overlap. While Universes that used ERI4 accounted for more than half of the total number of findings that were significant at the $p <= .05$ level in the *Extinction Retention Multiverse* (22 out of 38; 58%), there were no Universes that used ERI3 and found a significant result. This finding is consistent with prior evidence that, in addition to previously noted differences in the mechanisms measured by different types of ERI (see Lonsdorf, Merz, et al., 2019 for detailed review), a high number of methodological choices can yield variation in statistical findings through noise alone (see Simmons et al., 2011 for detailed discussion). We recommend that future fear learning studies aim to build on previous work (e.g., Lonsdorf, Merz, et al., 2019) to determine the degree to which variance in findings

across ERIs reflects statistical noise versus differences in specific clinically meaningful mechanisms of fear learning. In light of the multitude of effects and hypotheses to test simultaneously in the clinical literature, we also recommend following the trajectory of the genetics field by using very large samples combined with hypothesis-blind approaches in order to uncover robust effects (Dick et al., 2018; Duncan et al., 2019). Only once robust bases are formed, can we implement more advanced analyses such as deep phenotyping - an approach yielding promising advances in genetics (Dick et al., 2018; Duncan et al., 2019). This approach to fear learning may hold promise to inform future clinical breakthroughs due to the clear importance of fear learning as a clinically relevant mechanism, the abundance of extant archival data, and ongoing efforts to collect larger samples.

Although sex differences are of considerable interest in PTSD given the more than two-fold higher prevalence in women (Roberts et al., 2020), this study cannot reliably disentangle the effects of Sample Sex from the decrease in statistical power that comes with excluding a large portion of the sample. Similarly, our finding that nearly all results showing (non-significant) deficit in PTSD used only male participants is consistent with both prior evidence that fear extinction retention deficit in PTSD could be specific to males (Shvil et al., 2014), and also with previous findings that smaller samples can lead to unreliable effect size estimates (Button et al., 2013). Thus, larger studies are needed to directly examine sample sex composition as a potential moderator of fear extinction retention findings in PTSD and in other psychological disorders that are more prevalent in women (Kalin, 2020), while accounting for methodological factors.

Our finding of statistically significant evidence for fear acquisition (CS+ > CS−) across all 180 methodological combinations in our *Acquisition Multiverse* converges with findings from previous multiverse analyses that examined different methodologies in healthy samples. Specifically, previous multiverse analyses of fear conditioning found fear acquisition across most combinations of data reduction and analysis decisions (Lonsdorf et al., 2022) or across all combinations of SCR quantification decisions (Kuhn et al., 2022; Sjouwerman et al., 2022), and across multiple samples of healthy adults (Kuhn et al., 2022; Lonsdorf et al., 2022).[4] We extend those observations to a trauma-exposed clinical sample and to 180 combinations of 5 inconsistent methodologies that were derived from a review of the PTSD fear extinction retention literature and are relevant to the clinical fear learning literature more broadly.

[4]As described in footnote 1, the study by Kuhn et al. (2022) used a "manyverse analysis."

A comparison of our *Acquisition Multiverse* and *Extinction Retention Multiverse* results demonstrates that between-groups clinical effects may be less consistent in detecting a signal, less precise on average, and less consistent with respect to effect size than the within-subjects effect of acquisition. Specifically, lower precision in the *Extinction Retention Multiverse* as measured by wider 95% CIs is indicative of a greater average margin of error. Thus, in addition to having less consistency across methodologies, the point estimate of any given analysis is less reliable. Further, the *Extinction Retention Multiverse* was less consistent across all indices of consistency. Contrary to the direction of the statistically significant findings identified in our literature review, 15.1% of *Extinction Retention Multiverse* analyses found higher fear extinction retention in PTSD, while zero found decreased fear extinction retention in PTSD. These findings are consistent with the fact that within-subjects designs are inherently more powerful than between-subjects designs (Cote et al., 2021), and they reinforce the need to increase the statistical power of clinical group comparison studies of fear learning (Pöhlchen et al., 2020). Further, they add to previous evidence suggesting that meta-analysis alone may not be sufficient to address the low statistical power that is common in the clinical fear learning literature (e.g., Pöhlchen et al., 2020). Each of the studies included in our review interpreted their findings as providing at least partial support for the presence of a fear extinction retention deficit in PTSD (Garfinkel et al., 2014; Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008, 2009; Pineles et al., 2016; Shvil et al., 2014). However, the lack of consistent methodologies across these studies coupled with our *Extinction Retention Multiverse* results suggest that the size and direction of the estimated effect may be highly methodologically dependent, an observation that could not be derived from a meta-analysis that would have aggregated previously reported effects. Thus, our study provides additional support for the recommendation that a demonstration of methodological robustness should be a pre-requisite for performing meta-analysis (LeBel et al., 2018) and demonstrates that it may be particularly important to apply such an approach to the clinical fear learning literature specifically.

The higher impact of CAPS Exclusion in the *Acquisition Multiverse* relative to the *Extinction Retention Multiverse* suggests that symptom severity exclusions may sometimes have unanticipated effects on results. A common rationale for applying a symptom severity exclusion is to clearly separate the groups to be compared (Barker et al., 2015). In the case of trauma participants, exclusion of TENC with a CAPS score > 19 and of PTSD participants with a CAPS score < 50 is meant to minimize conceptual overlap

between the groups (i.e., remove subthreshold PTSD from the TENC group, and remove individuals in recovery or remission from the PTSD group) (Helpman et al., 2016). This approach is expected to impact the between-group comparison (PTSD versus TENC) but not the within-subjects effect (CS+ versus CS−) that are tested in the sample. However, in our data, CAPS Exclusion was the most impactful methodology in the *Acquisition Multiverse* and the second *least* impactful methodology in the *Extinction Retention Multiverse*, such that it had a more than 10-fold higher impact in the former. This finding is unlikely due to statistical power alone because Sample Sex has a larger impact on the sample size but less than one-fourth the impact on *Acquisition Multiverse* findings compared to CAPS Exclusion. Consistent with prior work showing that addition or removal of participants can have unintended consequences beyond statistical power (Schweinsberg et al., 2021; Simmons et al., 2011), we extend those findings to a type of methodological decision that is common in clinical studies (Barker et al., 2015).

An important limitation of this study is that our novel multiverse analysis was conducted in a single sample that was only well-powered to detect large effects. Although we did not find evidence of a fear extinction retention deficit in PTSD, it remains possible that a small-to-medium effect could be detected by applying these same methodologies to a larger sample (Lakens & Etz, 2017). However, it should be noted that we took steps to address this concern. Specifically, we used a larger sample ($n = 61$) than most previous studies of fear extinction retention in PTSD (Garfinkel et al., 2014; Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008, 2009; Orr et al., 2006; Pineles et al., 2016), comparable to the largest previously published report (Shvil et al., 2014). Further, in line with prior related studies that used one-tailed tests (Milad et al., 2008; Shvil et al., 2014), we conducted a supplemental multiverse analysis with one-tailed analyses, which doubled our statistical power yet still revealed no evidence of a fear extinction retention deficit in PTSD. However, our limited power coupled with our use of a single sample means that our findings on the relative impact of specific methodologies may not generalize to other samples. For example, although SCR Non-responder Exclusion had a relatively modest impact on findings in the *Extinction Retention Multiverse*, one study included in our review excluded more than twice as many of their participants (37%) (Pineles et al., 2016) as our study (16.4%) when using the same SCR Non-responder Exclusion criteria. Further, previous multiverse studies have found that the specific methodological combinations that maximize the effect size differ across samples (Kuhn et al., 2022; Lonsdorf et al., 2022). Thus, methodologies that had a small impact in our data may still be of high importance in other studies.

An additional noteworthy limitation is that our review and multiverse excluded methodologies that were used in prior investigations but were not feasible to apply to our data. Thus, it is possible that methodologies used in previous studies but not included in our *Extinction Retention Multiverse*, or additional methodologies that have yet to be applied to the study of fear extinction retention in PTSD, could identify consistent evidence of deficient fear extinction retention in PTSD. For example, differences in the type of fear learning task used and in SCR quantification methods differed in the reviewed literature but were beyond the scope of the present study (see Supplement for details). However, it is also noteworthy that the current study and 6 of the 7 studies included in our literature review (Garfinkel et al., 2014; Helpman et al., 2016; Marin et al., 2016; Milad et al., 2008, 2009; Shvil et al., 2014) used a version of the same fear task developed by Milad et al. (2005). The fact that most studies used a similar fear learning task is a strength of our review and multiverse analysis because it reduces variability attributable to task differences and increases our ability to detect effects of other variables on results. Similarly, all reviewed studies, and the present study, used a baseline correction approach for SCR quantification. Still, even small differences in those parameters have previously been found to impact results in the general fear learning literature (Fullana et al., 2020; Sjouwerman et al., 2022) and may interact with SCR Non-responder Exclusion (Lonsdorf, Klingelhöfer-Jens, et al., 2019). Thus, we recommend that future studies investigate their impact on fear learning findings in clinical samples. Further, SCR is only one of several measures to study fear extinction retention and other methods such as fear potentiated startle, subjective fear, and fMRI might prove more sensitive to detecting clinical group differences (Glover et al., 2011). However, methodological heterogeneity has also been observed across fear learning studies using each of those measures, and may impact findings (Lonsdorf & Merz, 2017). Further, converging evidence from the general neuroimaging literature suggests a critical need for studies to examine the impact of methodological variation on findings, and multiverse analysis has been suggested as a potential solution (for review see Niso et al., 2022). Thus, future studies using neuroimaging measures may also benefit from application of multiverse analysis to investigate the impact of methodological heterogeneity on clinical fear learning findings. Finally, although our study focuses on the conventional DSM-based diagnosis of PTSD, PTSD is clinically and biologically heterogeneous (Galatzer-Levy et al., 2017; Lewis et al., 2020; Liberzon, 2018). Thus, fear extinction retention deficits may exist in only a yet to be confirmed subset of PTSD patients (Liberzon, 2018; Lonsdorf & Merz, 2017), and we recommend that future studies apply multiverse analysis

to the exploration of fear extinction retention in theory-based (e.g., hypo versus hyper reactive (Lang et al., 2016)) or statistically-derived (Galatzer-Levy et al., 2017; Lewis et al., 2021) PTSD subgroups.

Although our review was limited in the scope of methodological parameters included, a wider scope would not have changed its main conclusions. Specifically, if we included all methodologies in our review, the conclusion that prior studies were methodologically heterogeneous would not have changed; in fact, we would have found more heterogeneity. Further, although we excluded three analyses from our review, those findings were also mixed and would not have changed the conclusion of inconsistent findings across prior investigations. Specifically, one excluded analysis found a significant deficit in PTSD versus healthy non-trauma-exposed co-twins (Milad et al., 2008); a second analysis found no difference between PTSD and healthy controls or participants with a fear-related disorder other than PTSD (Pöhlchen et al., 2020); and a third found no difference between PTSD and TENC after a seven-day paradigm in a study that also used a pharmacological manipulation (Orr et al., 2006).

In summary, we conducted a systematic literature review coupled with two Multiverse Analyses on a new data set to describe the effects of methodological variation on results from analyses of within-subject fear acquisition and group differences in fear extinction retention in PTSD. The hypothesis of acquisition (CS+ > CS−) was consistently supported across all analyses in the review and *Acquisition Multiverse*. By comparison, most analyses in our review failed to find evidence of deficient fear extinction retention in PTSD and our *Extinction Retention Multiverse* did not have any significant findings in the hypothesized direction. Further, our review found that between-study methodology was heterogeneous in this literature and our multiverse analyses demonstrated that the identified methodological differences had a greater overall impact on group comparisons of fear extinction retention. We echo previous statements that findings from single analysis studies, as well as meta-analytic studies that combine effects across single analysis studies, may be misleading (Patel et al., 2015; Simonsohn et al., 2020; Steegen et al., 2016) and we demonstrate that this extends to clinical group comparisons of fear learning. Finally, we demonstrate that multiverse analysis and Boba Sensitivity Analysis are valuable tools to quantify the impact of specific methodologies on results in clinical studies of psychophysiological fear learning. By clarifying the sources of conflicting results, multiverse analysis can help improve statistical inferences in clinical fear learning research.

AUTHOR CONTRIBUTIONS

Michael W. Lewis: Conceptualization; data curation; formal analysis; methodology; software; validation; visualization; writing – original draft; writing – review and editing. **Daniel E. Bradford:** Formal analysis; software; supervision; validation; visualization; writing – review and editing. **Edward F. Pace-Schott:** Data curation; methodology; resources; writing – review and editing. **Scott L. Rauch:** Investigation; methodology; supervision; writing – review and editing. **Isabelle M. Rosso:** Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision; writing – original draft; writing – review and editing.

ACKNOWLEDGMENTS

We thank our research participants, the research assistants who helped with data collection, and Dr. Yang Liu for her technical support.

FUNDING INFORMATION

A preliminary report of these data was presented as a poster presentation at the 2021 Society of Biological Psychiatry Annual Meeting. No portion of these data has been previously published. This research was supported by the National Institute of Mental Health (5R01MH096987[IMR]). IMR was also partially supported by R01MH120400 (IMR) and P50MH115874 (Carlezon, William, A.; Ressler, Kerry, J.). MWL was partially supported by a McLean Rappaport Mental Health Research Scholar Award. MWL and SLR were partially supported by P50MH115874 (Carlezon, Ressler). EFP was partially supported by 1R21MH121832 (EPS), R21MH128619 (EPS coPI) and Praxis Precision Medicines. SLR is employed by Mass General Brigham/McLean Hospital; he is paid as secretary of Society of Biological Psychiatry, and for Board service to Mindpath Health/Community Psychiatry; he has served as a volunteer member of the Board for Anxiety & Depression Association of America, and The National Network of Depression Centers; he has received royalties from Oxford University Press and American Psychiatric Publishing Inc; he has received research funding from NIMH. The study was approved by the Institutional Review Board of McLean Hospital and the Partners Human Research Committee (PHRC), IRB 2012P001237.

CONFLICT OF INTEREST STATEMENT

We have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Data and analysis code are openly available at Open Science Framework (https://osf.io/rybae/).

ORCID

Michael W. Lewis https://orcid.org/0000-0002-3903-9055
Daniel E. Bradford https://orcid.org/0000-0003-0920-6964
Edward F. Pace-Schott https://orcid.org/0000-0001-6311-7570
Scott L. Rauch https://orcid.org/0000-0001-7699-6754
Isabelle M. Rosso https://orcid.org/0000-0001-6988-3858

REFERENCES

Andero, R., & Ressler, K. (2012). Fear extinction and BDNF: Translating animal models of PTSD to the clinic. *Genes, Brain, and Behavior*, *11*(5), 503–512. https://doi.org/10.1111/j.1601-183X.2012.00801.x

Barker, C., Pistrang, N., & Elliott, R. (2015). *Research methods in clinical psychology: An introduction for students and practitioners*. John Wiley & Sons.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, *8*(1), 75–90. https://doi.org/10.1007/BF02105408

Boroumand, A., Ghose, S., Patel, M., Hassan, H., Lucia, B., Ausavarungnirun, R., Hsieh, K., Hajinazar, N., Malladi, K. T., Zheng, H., & Mutlu, O. (2019). CoNDA: Efficient cache coherence support for near-data accelerators. *Proceedings of the 46th International Symposium on Computer Architecture*, 629–642. https://doi.org/10.1145/3307650.3322266

Bradford, D. E., DeFalco, A., Perkins, E., Carbajal, I., Kwasa, J., Goodman, F., Jackson, F., Richardson, L. N. S., Woodley, N., Neuberger, L., Sandoval, J. A., Huang, H. J., & Joyner, K. (2022). Whose signals are we amplifying? Towards a more equitable clinical psychophysiology. *Clinical Psychological Science*. Advance online publication. https://doi.org/10.31234/osf.io/c2naf

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cote, L. R., Gordon, R., Randell, C. E., Schmitt, J., & Marvin, H. (2021). *Introduction to statistics in the psychological sciences*. Open Educational Resources Collection. https://umsystem.pressbooks.pub/isps/

Del Giudice, M., & Gangestad, S. W. (2021). A traveler's guide to the multiverse: Promises, pitfalls, and a framework for the evaluation of analytic decisions. *Advances in Methods and Practices in Psychological Science*, *4*(1). https://doi.org/10.1177/2515245920954925

Dick, D. M., Barr, P. B., Cho, S. B., Cooke, M. E., Kuo, S. I. C., Lewis, T. J., Neale, Z., Salvatore, J. E., Savage, J., & Su, J. (2018). Post-GWAS in psychiatric genetics: A developmental perspective on the "other" next steps. *Genes, Brain and Behavior*, *17*(3), e12447.

Duncan, L. E., Ostacher, M., & Ballon, J. (2019). How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology*, *44*(9), 1518–1523. https://doi.org/10.1038/s41386-019-0389-5

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology*, *18*(3), 232–239. https://doi.org/10.1111/j.1469-8986.1981.tb03024.x

Fullana, M., Dunsmoor, J., Schruers, K., Savage, H., Bach, D., & Harrison, B. (2020). Human fear conditioning: From neuroscience to the clinic. *Behaviour Research and Therapy*, *124*, e103528. https://doi.org/10.1016/j.brat.2019.103528

Galatzer-Levy, I. R., Andero, R., Sawamura, T., Jovanovic, T., Papini, S., Ressler, K. J., & Norrholm, S. D. (2017). A cross species study of heterogeneity in fear extinction learning in relation to FKBP5 variation and expression: Implications for the acute treatment of posttraumatic stress disorder. *Neuropharmacology*, *116*, 188–195. https://doi.org/10.1016/j.neuropharm.2016.12.023

Garfinkel, S. N., Abelson, J. L., King, A. P., Sripada, R. K., Wang, X., Gaines, L. M., & Liberzon, I. (2014). Impaired contextual modulation of memories in PTSD: An fMRI and psychophysiological study of extinction retention and fear renewal. *The Journal of Neuroscience*, *34*(40), 13435–13443. https://doi.org/10.1523/JNEUROSCI.4287-13.2014

Glover, E. M., Phifer, J. E., Crain, D. F., Norrholm, S. D., Davis, M., Bradley, B., Ressler, K. J., & Jovanovic, T. (2011). Tools for translational neuroscience: PTSD is associated with heightened fear responses using acoustic startle but not skin conductance measures. *Depression and Anxiety*, *28*(12), 1058–1066. https://doi.org/10.1002/da.20880

Gray, M. J., Litz, B. T., Hsu, J. L., & Lombardo, T. W. (2004). Psychometric properties of the life events checklist. *Assessment*, *11*(4), 330–341. https://doi.org/10.1177/1073191104269954

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Lawrence Erlbaum Associates Publishers. https://psycnet.apa.org/record/2005-04135-000

Helpman, L., Marin, M.-F., Papini, S., Zhu, X., Sullivan, G. M., Schneier, F., Neria, M., Shvil, E., Malaga Aragon, M. J., Markowitz, J. C., Lindquist, M. A., Wager, T., Milad, M., & Neria, Y. (2016). Neural changes in extinction recall following prolonged exposure treatment for PTSD: A longitudinal fMRI study. *NeuroImage: Clinical*, *12*, 715–723. https://doi.org/10.1016/j.nicl.2016.10.007

Kalin, N. H. (2020). The critical relationship between anxiety and depression. *The American Journal of Psychiatry*, *177*(5), 365–367. https://doi.org/10.1176/appi.ajp.2020.20030305

Kredlow, M. A., Pineles, S. L., Inslicht, S. S., Marin, M. F., 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

Kuhn, M., Gerlicher, A. M. V., & Lonsdorf, T. B. (2022). Navigating the manyverse of skin conductance response quantification approaches - a direct comparison of trough-to-peak, baseline correction, and model-based approaches in Ledalab and PsPM. *Psychophysiology*, *59*(9), e14058. https://doi.org/10.1111/psyp.14058

Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science*, *8*(8), 875–881. https://doi.org/10.1177/1948550617693058

Lang, P. J., McTeague, L. M., & Bradley, M. M. (2016). RDoC, DSM, and the reflex physiology of fear: A biodimensional analysis of

the anxiety disorders spectrum. *Psychophysiology*, *53*(3), 336–347. https://doi.org/10.1111/psyp.12462

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*(3), 389–402. https://doi.org/10.1177/2515245918787489

Lewis, M. W., Friedman, B. H., & Jones, R. T. (2021). Increases and decreases in fear potentiated startle during fear acquisition: A latent class growth analysis. *Psychology & Neuroscience*, *14*(4), 396–412. https://doi.org/10.1037/pne0000266

Lewis, M. W., Jones, R. T., & Davis, M. T. (2020). Exploring the impact of trauma type and extent of exposure on posttraumatic alterations in 5-HT1A expression. *Translational Psychiatry*, *10*(1), e237. https://doi.org/10.1038/s41398-020-00915-1

Liberzon, I. (2018). Searching for intermediate phenotypes of posttraumatic stress disorder. *Biological Psychiatry*, *83*(10), 797–799. https://doi.org/10.1016/j.biopsych.2017.06.005

Liu, Y., Kale, A., Althoff, T., & Heer, J. (2021). Boba: Authoring and visualizing multiverse analyses. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 1753–1763. https://doi.org/10.1109/TVCG.2020.3028985

Lonsdorf, T. B., Gerlicher, A., Klingelhöfer-Jens, M., & Krypotos, A.-M. (2022). Multiverse analyses in fear conditioning research. *Behaviour Research and Therapy*, *153*, 104072. https://doi.org/10.1016/j.brat.2022.104072

Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., Jentsch, V. L., Drexler, S. M., & Merz, C. J. (2019). Navigating the garden of forking paths for data exclusions in fear conditioning research. *eLife*, *8*, e52465. https://doi.org/10.7554/eLife.52465

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans-biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience and Biobehavioral Reviews*, *80*, 703–728. https://doi.org/10.1016/j.neubiorev.2017.07.007

Lonsdorf, T. B., Merz, C. J., & Fullana, M. A. (2019). Fear extinction retention: Is it what we think it is? *Biological Psychiatry*, *85*(12), 1074–1082. https://doi.org/10.1016/j.biopsych.2019.02.011

Marin, M.-F., Song, H., VanElzakker, M. B., Staples-Bradley, L. K., Linnman, C., Pace-Schott, E. F., Lasko, N. B., Shin, L. M., & Milad, M. R. (2016). Association of resting metabolism in the fear neural network with extinction recall activations and clinical measures in trauma-exposed individuals. *The American Journal of Psychiatry*, *173*(9), 930–938. https://doi.org/10.1176/appi.ajp.2015.14111460

Milad, M. R., Furtak, S. C., Greenberg, J. L., Keshaviah, A., Jooyeon, J. I., Falkenstein, M. J., Jenike, M., Rauch, S. L., & Wilhelm, S. (2013). Deficits in conditioned fear extinction in obsessive-compulsive disorder and neurobiological changes in the fear circuit. *JAMA Psychiatry*, *70*(6), 608–618. https://doi.org/10.1001/jamapsychiatry.2013.914

Milad, M. R., Orr, S. P., Lasko, N. B., Chang, Y., Rauch, S. L., & Pitman, R. K. (2008). Presence and acquired origin of reduced recall for fear extinction in PTSD: Results of a twin study. *Journal of Psychiatric Research*, *42*(7), 515–520. https://doi.org/10.1016/j.jpsychires.2008.01.017

Milad, M. R., Orr, S. P., Pitman, R. K., & Rauch, S. L. (2005). Context modulation of memory for fear extinction in humans. *Psychophysiology*, *42*(4), 456–464. https://doi.org/10.1111/j.1469-8986.2005.00302.x

Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., Zeidan, M. A., Handwerger, K., Orr, S. P., & Rauch, S. L. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biological Psychiatry*, *66*(12), 1075–1082. https://doi.org/10.1016/j.biopsych.2009.06.026

Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological Psychiatry*, *62*(5), 446–454. https://doi.org/10.1016/j.biopsych.2006.10.011

Morís Fernández, L., & Vadillo, M. A. (2020). Flexibility in reaction time analysis: Many roads to a false positive? *Royal Society Open Science*, *7*(2), 190831. https://doi.org/10.1098/rsos.190831

Niso, G., Botvinik-Nezer, R., Appelhoff, S., De La Vega, A., Esteban, O., Etzel, J. A., Finc, K., Ganz, M., Gau, R., Halchenko, Y. O., Herholz, P., Karakuzu, A., Keator, D. B., Markiewicz, C. J., Maumet, C., Pernet, C. R., Pestilli, F., Queder, N., Schmitt, T., … Rieger, J. W. (2022). Open and reproducible neuroimaging: From study inception to publication. *NeuroImage*, *263*, 119623. https://doi.org/10.1016/j.neuroimage.2022.119623

Orr, S. P., Milad, M. R., Metzger, L. J., Lasko, N. B., Gilbertson, M. W., & Pitman, R. K. (2006). Effects of beta blockade, PTSD diagnosis, and explicit threat on the extinction and retention of an aversively conditioned response. *Biological Psychology*, *73*(3), 262–271. https://doi.org/10.1016/j.biopsycho.2006.05.001

Patel, C. J., Burford, B., & Ioannidis, J. P. A. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, *68*(9), 1046–1058. https://doi.org/10.1016/j.jclinepi.2015.05.029

Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M., & Vrgoc, D. (2016). Foundations of JSON schema. In *Proceedings of the 25th international conference on world wide web* (pp. 263–273). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland. https://doi.org/10.1145/2872427.2883029

Pineles, S. L., Nillni, Y. I., King, M. W., Patton, S. C., Bauer, M. R., Mostoufi, S. M., Gerber, M. R., Hauger, R., Resick, P. A., Rasmusson, A. M., & Orr, S. P. (2016). Extinction retention and the menstrual cycle: Different associations for women with posttraumatic stress disorder. *Journal of Abnormal Psychology*, *125*(3), 349–355. https://doi.org/10.1037/abn0000138

Pöhlchen, D., Leuchs, L., Binder, F. P., Blaskovich, B., Nantawisarakul, T., Topalidis, P., Brückl, T. M., Norrholm, S. D., Jovanovic, T., BeCOME working group, & Spoormaker, V. I. (2020). No robust differences in fear conditioning between patients with fear-related disorders and healthy controls. *Behaviour Research and Therapy*, *129*, 103610. https://doi.org/10.1016/j.brat.2020.103610

Ponniah, K., & Hollon, S. D. (2009). Empirically supported psychological treatments for adult acute stress disorder and posttraumatic stress disorder: A review. *Depression and Anxiety*, *26*(12), 1086–1109. https://doi.org/10.1002/da.20635

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rauch, S. A. M., Eftekhari, A., & Ruzek, J. I. (2012). Review of exposure therapy: A gold standard for PTSD treatment. *Journal*

of Rehabilitation Research and Development, *49*(5), 679–687. https://doi.org/10.1682/jrrd.2011.08.0152

Roberts, A. L., Kubzansky, L. D., Chibnik, L. B., Rimm, E. B., & Koenen, K. C. (2020). Association of posttraumatic stress and depressive symptoms with mortality in women. *JAMA Network Open*, *3*(12), e2027935. https://doi.org/10.1001/jamanetworkopen.2020.27935

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., Liu, Y., Althoff, T., Heer, J., Kale, A., Mohamed, Z., Amireh, H., Prasad, V. V., Bernstein, A., Robinson, E., Snellman, K., Sommer, S. A., Otner, S. M. G., Robinson, D., … Uhlmann, E. L. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes*, *165*, 228–249. https://doi.org/10.1016/j.obhdp.2021.02.003

Shvil, E., Sullivan, G. M., Schafer, S., Markowitz, J. C., Campeas, M., Wager, T. D., Milad, M. R., & Neria, Y. (2014). Sex differences in extinction recall in posttraumatic stress disorder: A pilot fMRI study. *Neurobiology of Learning and Memory*, *113*, 101–108. https://doi.org/10.1016/j.nlm.2014.02.003

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Sjouwerman, R., Illius, S., Kuhn, M., & Lonsdorf, T. B. (2022). A data multiverse analysis investigating non-model based SCR quantification approaches. *Psychophysiology*, *59*(12), e14130. https://doi.org/10.1111/psyp.14130

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual Review of Clinical Psychology*, *15*, 579–604. https://doi.org/10.1146/annurev-clinpsy-050718-095710

Vervliet, B., & Boddez, Y. (2020). Memories of 100 years of human fear conditioning research and expectations for its future. *Behaviour Research and Therapy*, *135*, 103–123. https://doi.org/10.1016/j.brat.2020.103732

Webb, E. K., Etter, J. A., & Kwasa, J. A. (2022). Addressing racial and phenotypic bias in human neuroscience methods. *Nature Neuroscience*, *25*(4), 410–414. https://doi.org/10.1038/s41593-022-01046-0

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1

How to cite this article: Lewis, M W., Bradford, D E., Pace-Schott, E F., Rauch, S L., & Rosso, I M. (2023). Multiverse analyses of fear acquisition and extinction retention in posttraumatic stress disorder. *Psychophysiology*, *60*, e14265. https://doi.org/10.1111/psyp.14265

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.