Fear extinction retention – is it what we think it is?

Running title: Extinction retention

Tina B. Lonsdorf[1] *, Christian J. Merz[2] & Miquel A. Fullana[3]

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

[2] Ruhr-University Bochum, Institute of Cognitive Neuroscience, Department of Cognitive Psychology, Bochum, Germany.

[3] Institute of Neurosciences, Hospital Clinic, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Barcelona, Spain

* corresponding author: Institute for Systems Neuroscience, University Medical Center Hamburg Eppendorf, Martinistraße 52, 20246 Hamburg, Germany, t.lonsdorf@uke.de, Phone: +49 (0) 40 7410 – 55769, Telefax: +49 (0) 40 7410 – 59955

Number of words (abstract): 151

Number of words: (main body of text): 3140

Number of Figures: 1

Number of Tables: 1

Number of Supplementary Material: 1

Abstract

There has been an explosion of research on fear extinction in humans in the past two decades. This has generated major insights but also brought a new goal into focus: how to maintain extinction memory over time (i.e., extinction retention). We argue that there are still important conceptual and procedural challenges in human fear extinction research that hamper advancement in the field. We use extinction retention and the 'extinction retention index (ERI)' to exemplarily illustrate these challenges. Our systematic literature search identified 16 different operationalizations of the ERI. Correlation coefficients between these different operationalizations as well as with measures of fear/anxiety show a wide range of variability in four independent datasets with similar findings across datasets. Our results suggest that there is an urgent need for standardization in the field. We discuss the conceptual and empirical implications of these results and provide specific recommendations for future work.

Keywords: extinction retention index, extinction recall index, retrieval index, systematic literature search, fear conditioning, meta-research

Introduction

In 2006, Anderson and Insel stated that "The development of new approaches to anxiety disorders based on the neurobiology of fear extinction represents perhaps the best current opportunity for translating neuroscience discoveries into clinical applications[…]" (cf. 1; page 319). Since then, there has been an enormous growth in fear extinction research (e.g., 2–6) which continues two decades later (7, 8). This generated major insights into extinction mechanisms (7, 9–12) but has also brought a new goal into focus: "The current challenge however is not how to *achieve* fear reduction [i.e., extinction], but rather to *maintain* it over time [i.e. extinction retention]" (cf. 8). Here, we argue that despite decades of research, there are conceptual and procedural challenges that urgently need to be addressed for experimental research on extinction retention to successfully translate into clinical applications.

Extinction and extinction retention: Conceptual challenges

Extinction has been typically investigated in 'fear conditioning experiments'[1] (13): *Acquisition* of conditioned fear is achieved by presenting an initially neutral stimulus (conditional stimulus, CS+) paired with an aversive event (unconditioned stimulus, US), which generates a fear (CS+/US) memory (a procedure termed *fear acquisition training)*. While rodent work typically includes only a CS+ (single-cue protocols), human work typically includes a second stimulus (CS-) not followed by the US (differential protocols). Importantly, conditioned responding is quantified as differential responding ((CS+)-(CS)) in differential protocols.

When the CS+ is no longer followed by the US for a sufficient number of trials, the CR gradually disappears (a procedure called *extinction training)*. The contemporary view is that the original conditioned fear memory is not erased, but inhibited by a competing extinction memory (14). Upon presentation of the CS at a later time (*i.e., retention test*), the dominance of one of these memories over the other determines whether fear is expressed (*fear retention*) or not (*extinction retention*). Experimental protocols designed to investigate extinction retention (e.g., 15) sometimes include two

different CS+ types during fear acquisition, with only one being subsequently extinguished (CS+e) while the second is not presented during extinction training (CS+u; unextinguished). Methods in human 'fear conditioning' are heterogeneous and even subtle procedural variations impact on learning *processes* (discussed in (13)). The term '*extinction retention'* has been used to refer to different procedural scenarios (13). Typically, a test-phase following after (e.g., 24h) extinction learning is referred to as "extinction retention phase". However, strictly speaking, this is only appropriate when contextual manipulations that likely trigger dominance of extinction over fear memory are employed – such as the test phase taking place in the extinction learning context (i.e., $A_{acq}B_{ext}B_{test}$). In the absence of such manipulations (e.g., $A_{acq}A_{ext}A_{test}$ paradigm), there is no reason to believe that the extinction memory is more likely to be retrieved than the fear memory. This is illustrated by 'spontaneous recovery' often used to refer to the very same procedure as 'extinction retention' (13). Here, as noted previously (5, 13), the distinction between *procedure and process* is of utmost importance. More precisely, we argue that a test phase (i.e., procedure) following extinction should be referred a-theoretically as 'retention test' (13) during which the re-occurrence of conditioned responding or its absence may be *observed* or *hypothesized*. Accordingly, the *processes* underlying the observed results should be referred to as 'return of fear' or 'extinction retention' respectively.

Extinction and extinction retention: Procedural challenges

The operationalization of extinction and extinction retention also varies widely (see (13)), which we illustrate here by using the 'extinction retention index' (ERI) as an example:

The ERI - as employed in rodent work using freezing (e.g., 16)- was introduced to the human field using skin conductance responses (SCR) (17–19) as a cross-species translational tool. The ERI followed on the idea that the strength of the responding during a retention test can be expressed as the percentage of the strength of such responding during fear acquisition (i.e., "how much fear comes back of the fear acquired?"). For illustration, consider two individuals, one ("X") showing higher maximal SCR CS+

responses (1µS) than the other ("Y") during acquisition training (0.5µS). Subsequently, both individuals undergo extinction training. During a later retention test, both individuals display the same amount of CS+ responses (i.e., 0.5µS). Consequently, X´s extinction retention would be considered more efficient as compared to Y's, as he/she shows less CS+ responses at the retention test with respect to the CS+ responses during acquisition training (based on an example provided by M.R. Milad).

Since its introduction in humans, the ERI has been widely employed – in particular for SCRs - and is assumed to represent a standardized index that supports both comparability and replicability of findings. However, our systematic literature search identified 16 different calculations of the ERI using SCR. To illustrate the potential impact of this subtle - but often unrecognized – heterogeneity, we have re-analyzed four datasets to calculate the magnitude of the correlations between these different ERIs. Our results challenge the conceptual and empirical rationale for the ERI. Finally, we provide recommendations for future work.

Methods and Materials

We conducted a systematic literature search to identify peer-reviewed studies published until October 2018 in which an "extinction retention index" (ERI) was calculated using SCR in humans (see Supplement for details). Subsequently, we used SCR data from a published study (dataset 1 with N=50 (20)) to (re)calculate the ERIs using the formulas identified by the literature search. In short, 50 healthy participants with moderate to strong fear of spiders underwent a two-day differential (CS+, CS-) paradigm (day 1: fear acquisition, immediate extinction; day 2: extinction retention; see Supplement for details). Finally, we calculated Spearman's rank coefficients between the different ERIs (see Table 1), since we were interested if the specific rank between participants changes across ERI versions. The ERIs were also (re)calculated in three additional datasets (datasets 2-4) all using a two day (i.e., immediate extinction (21)) or three day (i.e., delayed extinction (21, 22)) paradigm including two CS+s (CS+e, CS+u) and one CS- in healthy participants (see Supplementary material for methodological details and results). In addition, inspired by reviewer comments, correlations between measures of

fear/anxiety and the ERIs were calculated. In dataset 1, the Fear of Spiders Questionnaire (FSQ (23)) was used, while the State-Trait Anxiety Inventory (STAI (24)) was used in datasets 2-4. P-values were corrected for multiple comparisons using the Benjamini-Hochberg procedure (25) separately for cross-ERI correlations and correlations between the ERIs and the FSQ and STAI respectively.

Results

Heterogeneity in ERI calculation: We identified 16 different calculations of the ERI included in 37 separate studies (see **Table 1** and note that three studies (26–28) included two different ERI versions[2]) and a total 34% of studies using SCRs during a retention test employed 'an ERI'. In n=26 studies, the retention test took place in the extinction learning context (i.e. testing for extinction retention), while in n=11 studies, no contextual manipulation was applied.

The ERI calculations identified differed in a multitude of ways. *First*, responding during the retention test was operationalized as differential responding (i.e., difference between the CS+ and the CS-) in n=9 studies (henceforth "differential ERIs") and as responding to the CS+ only in n=28 studies (henceforth "non-differential ERIs", one study (27) used in addition a CS- based index). *Second*, the number of trials the ERI was based on ranged from one to five (one: n=4; two: n=19; three: n=1; four: n=13; five: n=2) – a wide range in light of rapidly occurring re-extinction due to non-reinforced CS presentations during retention test. *Third*, responding during the retention test was corrected for responding during acquisition (n=31 studies) or extinction (n=2) while also no correction was employed (n=4).

Fourth, of those 31 studies 'correcting responding for the strength of fear learning' (cf. 29), responding during acquisition training was operationalized as the maximum response to the CS+ (n=9), the CS+e (n=13), any CS+ (i.e., CS+e or CS+u; n=1), or any CS (n=1), the average of the two largest responses to the CS+ (n=3), or the differential response (maximum CS+/CS- difference; n=4). The maximum CS+ response during acquisition training however, may not be a good indicator of the 'strength of fear

learning'. For instance, in our data, the maximum SCR to a CS+ is most often observed to the very first CS+ (see **Figure 1A** and **1B**) which *precedes* the first US presentation and hence reflects rather arousal or orienting (30) than associative learning strength. In contrast, the maximum *differential* responding between a pair of CS+ and CS- presentations is typically observed at the very end of acquisition training (illustrated in our data in **Figure 1B**). Hence, the maximum differential responding during fear acquisition training is more likely to relate to associative learning processes, as it would be the case for maximum freezing to the CS+ in rodents (see **Figure 1B** right). Note however, that only a few studies have employed differential responding during acquisition training to calculate the ERI (see **Table 1**).

Correlations between ERIs: Correlations between the 16 identified ERIs in our dataset, ranged from 0.003 to (-)1 (see **Figure 1D**, note that the algebraic sign will be ignored henceforward as it does only reflect the interpretation as % fear recalled or % fear not recalled). Overall, non-differential and differential indices emerged as two 'distinct clusters' (with the exception of the single CS- based index 9b), with correlations ranging between .27 and 1 within non-differential ERIs and between .5 and .93 within differential ERIs. The correlations between differential and non-differential ERIs ranged between .19 and .61. Results of the additional datasets show a similar pattern of correlations (see Supplementary Figure 2).

Correlations between ERIs and FSQ ranged between (-).03 and (-).26 (again, ignoring the negative algebraic sign) and all correlations between FSQ and any of the ERIs were non-significant (see Figure 1D).

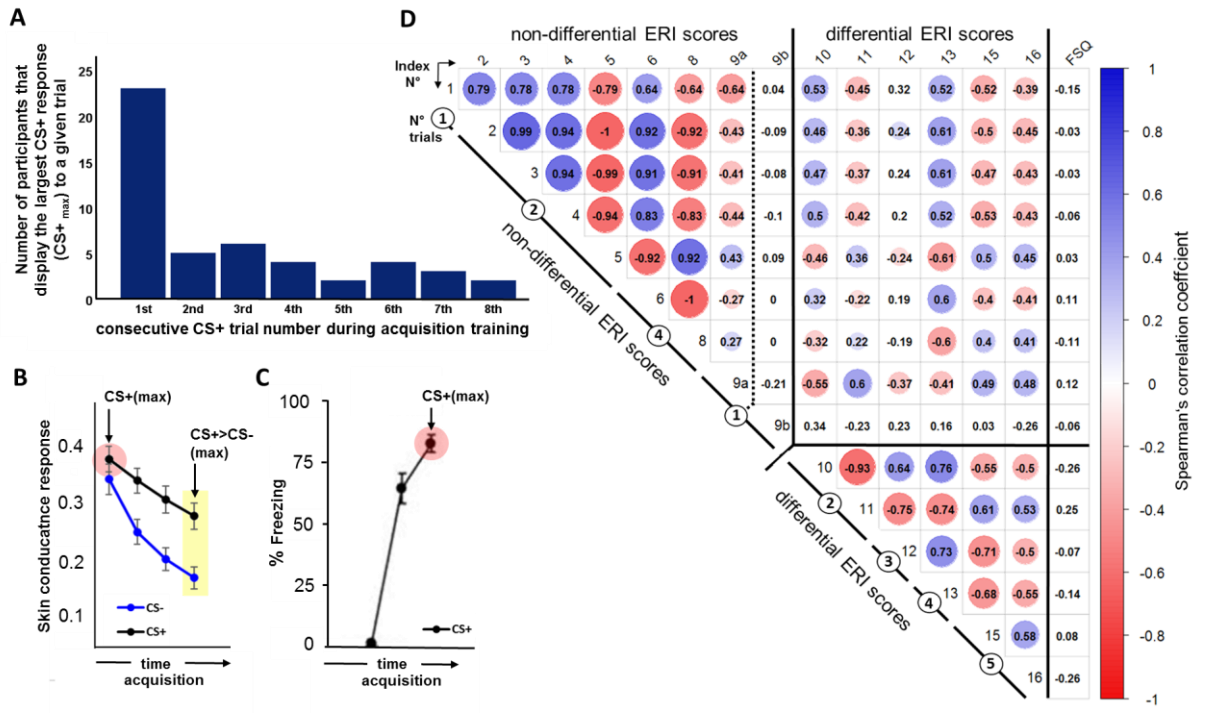


Figure 1 (A) Number of individuals in our dataset (from N=50) that displayed the maximum CS+ response to each of the eight CS+ trials during acquisition training. (**B**) Acquisition trials reflecting maximum responding to the CS+ (highlighted in red) as well as maximum differential responding (i.e., CS+>CS-, highlighted in yellow) in the present study in humans employing a differential conditioning protocol with 100% reinforcement (blocks of 2 trials shown; the outcome measure was skin conductance responding). **(C)** Acquisition trials reflecting maximum responding to the CS+ (highlighted in red) in a rodent study employing a single-cue conditioning protocol (blocks of two trials shown; the outcome measure was freezing) (figure modified after 16)[a]. **(D)** Correlation matrix (Spearman's r) between the different ERIs formulas as derived from our systematic literature search (as indicated in Table 1) as well as the FSQ as re-calculated based in our data (see above). Correlations are illustrated as a heat map (blue: significant positive correlation, red: significant negative correlation, white cell: non-significant correlation (i.e., $p>0.05$) using the 'corrplot' package in R. Correction for multiple comparisons was applied separately for the cross-ERI correlations and the correlation between the ERIs and the FSQ respectively.

Note that index 7 is not included in the correlation matrix as it is identical to index 6 when calculated in our dataset because the dataset used for calculations did not include a CS+e and CS+u but only a single CS+. Index 14 is not included here, as it is based on the difference between the CS+e and CS+u, which are not available in this dataset. We, however, refer to the supplementary material for results of additional datasets (dataset 2-4) that employ these two different CS+ types (CS+e and CS+u) as well as a partial reinforcement rate and immediate (dataset 3) and delayed extinction (dataset 2 and 4).

Note that the negative correlations between some of the ERIs (such as ERI N°5 and N°8 and N°9a) with the other non-differential ERIs (i.e., ERI N°1- N°4 and N° 6), result from the fact that the latter subtract the retention score (i.e., responding during retention divided by responding during acquisition) from 100 which yielding the percentage of 'fear not recovered (i.e., extinction retention)' whereas ERIs N°5, N°8 and N°9a reflect the percentage of 'fear recovered'. While the interpretation of the score is thus inverse, the sign of the correlation (i.e., positive or negative) is not of primary interest to our question and is hence ignored henceforward.

Errorbars show s.e.m.

Legend: ERI: Extinction retention index, CS+: conditioned stimulus +, CS-: non-conditioned stimulus, CS+e: extinguished CS+, CS+u: unextinguished CS+

[a] Figure modified with permission from the publisher and author. Copyright by Cold Spring Harbor Laboratory Press.

Discussion

Precision in concepts, methods and data analysis is key to science. By using the extinction retention index (ERI) as an example, we have illustrated the problem of –often unrecognized- heterogeneity in operationalization for fear extinction retention research in humans. Awareness to these matters is an important first step (31) towards more rigor in the field and successful translation into clinical applications.

First, we have highlighted that the term 'extinction retention' is often employed despite experimental designs not allowing to infer dominant recall of extinction memory (i.e., $A_{acq}A_{ext}A_{test}$ paradigm, see Table 1) which is misleading.

Second, from a *procedural* perspective, we show substantial variation in the calculation of an 'extinction retention index' with unsatisfactory correlations between the 16 different ERI versions across four datasets (20–22). We hence argue that the ERI, initially intended to be a cross-species translational measure, has evolved into a set of idiosyncratic 'formulas'. This may hamper replicability and advancement in the field (32, 33).

Third, from a *conceptual* perspective, we highlight below that none of the 16 different ERI formulas can be recommended as a good operationalization of the theoretical construct of 'extinction retention'.

Does the extinction recall index make sense from a conceptual perspective?

The rationale for the ERI is to express responding during a retention test as a percentage of responding during acquisition (29).

According to prevailing extinction theories (6, 14) however, whether fear will re-occur at this later test (i.e., return of fear) or not (i.e., recall of extinction) is determined by the dominance of the fear memory over the extinction memory (or vice versa) – hence on *both* the fear *and* the extinction memory. Thus, it is surprising that most ERIs have controlled for responding during acquisition training, whereas

control for extinction is very rare (and control for both has not been reported) – implying that extinction will be similarly efficient for all individuals. For instance, two individuals X and Y showing identical CS+ responding (0.5µS) during retention test after different amounts of CS+max responses (1µS versus 0.5µS) during acquisition training. Normalizing CS+ responses during the retention test for CS+max(acq) (i.e., index 1) would yield a 50% extinction retention for the X [i.e., 100-(100*1µS/0.5µS] but 0% for Y [i.e., 100-(100*0.5µS/0.5µS] and we would infer better extinction in X than in Y. Moreover, not only the strength but also the consolidation of fear and extinction memory acquisition are crucial for later retention. The major role of consolidation processes is illustrated by the fact that *within-session* extinction learning is not significantly correlated with *between-session* extinction learning (34) or performance at a later test in humans (35) or rodents (36–38). In our example, X and Y may show an identical amount of CS+max responding (0.5µS) during acquisition training but might undergo efficient or inefficient consolidation of fear memory respectively. When these individuals show different amounts of CS+ responding (1.0µS versus 0.5µS) during the retention test, the ERI (typically claiming to correct for acquisition performance) would however attribute these to the retention of extinction rather than possibly different levels of consolidation of fear memory.

In sum, we argue that the theoretical foundation of the ERI to express responding at a retention test as a fraction of responding during fear acquisition training, as employed in most ERIs, does not map well onto prevailing theories and empirical findings. In addition, none of the ERIs showed a consistent association to measures of fear/anxiety across datasets (i.e., FSQ, STAI). In fact, there was a consistency in the absence of such a relation.

Does the operationalization of the extinction recall index (ERI) make sense?

Here, we identified 16 different operationalizations of "the ERI", all intended to capture the same process (i.e., extinction retention) but empirically showing unsatisfactory correlations across four datasets. Importantly, although the four datasets used different procedures (e.g., immediate vs. delayed extinction training, CS+ vs. CS+e and CS+u) the pattern of correlations across ERIs is very similar

across datasets. This highlights the robustness and generalizability of our findings. Of note, the non-differential ERIs seem to be more related to each other than the differential ERIs, probably indicating that there is less variability in the former than in the latter (e.g., ERIs 10 and 13 control for extinction retention with acquisition data, whereas ERIs 12 and 14 do not).

Importantly, the ERI has been translated from rodent freezing (e.g., 16) to human work mainly using SCRs. Procedural differences between rodent and human work may however limit direct "translationability" of the ERI: Rodent work employs mostly single-cue designs (i.e., CS+ only), while human work employs almost exclusively differential designs (CS+ vs. CS-). Remarkably, despite differential designs, most ERIs employed in humans are *non-differential* (i.e., including CS+/ CS+e only; Table 1), which is problematic: First, the CS- was introduced to control for general responsivity and non-associative processes such as arousal or orienting (13), and conditioned responding is typically quantified as differential (i.e., CS+ vs. CS-) responding. As such, the "typical" ERI calculations (e.g. MEAN CS+ responding during recall)/(max CS+ responding during acquisition) may capture general arousal/orienting rather than associative processes. Second, CS+(max) responding during acquisition does not seem to reflect acquisition strength. As illustrated in Figure 1B, the maximum CS+ response during acquisition is most frequently observed to the very first CS+ presentation preceding the first CS+/US pairing and therefore reflect rather orienting (30). To control for potential effects of this orienting response during extinction retention, some authors have established that the first trial during retention is always a CS− and disregarded this first trial in the calculations of the ERI (cf. 27). Importantly, in freezing, the CS+(max) typically occurs at the end of acquisition, illustrating the challenges and limitations of direct cross-species translation.

Similarly, SCRs to the first CS+ at retention test may primarily reflect orienting and arousal when considered in isolation (i.e., without comparing to the CS-). As a consequence, non-differential ERIs cannot answer the question they intending to (i.e., 'how much of the acquired fear comes back').

Of note, while traditionally employed for SCRs, the ERI has been also expanded to other outcomes measures (i.e., FPS, ratings) recently (20, 39–42). Importantly, the conceptual problems we discuss in this work also apply to these other outcome measures. In addition, ERIs including correction for CS+max responding during acquisition are not widely applicable to functional magnetic resonance imaging (fMRI) data as single trial analyses are inherently difficult in fMRI. Consequently, studies using multiple outcome measures often employ an ERI for SCRs, but base their critical calculations for other outcome measures on different calculations, rendering the results not comparable.

It is also important to note that different operationalizations of the ERI tap into different clinically relevant mechanisms. Patients have been shown to display deficits particularly in extinction learning and safety signal (i.e., CS-) processing (23, 24) – both of which are not accounted for in the current ERI operationalizations – particularly in non-differential operationalizations.

In closing, we have exemplarily challenged both the conceptual foundations and procedural operationalization of 'extinction retention'. While a standardized way to quantify 'retention' in an interpretable way is highly desirable, the complexity of processes, aims and consequentially experimental designs in the field renders a simple 'gold standard' solution impractical (13). Recommendations that can be derived from our work include 1) preferring differential responding over isolated CS+ responding; 2) refraining from employing $CS+_{max}$ responses during acquisition training as a measure of associative learning; and 3) appreciating the relevance of fear and extinction memory strength and their respective consolidation, which implies that correcting for one of these factors but not for the others will likely introduce a bias.

Here, we provide conceptual and empirical arguments that speak against the employment of an "ERI", which leads to massive data reduction and hence interpretation problems. Rather than using an ERI, we suggest relying on within-session (i.e., retention test) *differential* responding rather than merely CS+ based responses. Furthermore, we suggest considering the dynamics over time (13, 43) and providing trial-by-trial data (whenever possible) for all stimuli and phases included in the experimental

design (i.e., CS+ or CS+e/CS+u and CS-) as well as for all outcome measures. Other general recommendations, such as justifying the exclusion of participants and demonstrating the invariance of the results regarding exclusions if employed (13, 44) as well as the use of hierarchical models over traditional ANOVAs (45), apply here as well. Yet, specific analysis choices that may depend on the specific design, such as the number of trials included/excluded, still need to be justified and reported in a transparent way.

Finally, raising awareness to the threat of (unrecognized) methodological and data analytical heterogeneity will hopefully 1) spark similar approaches in other subfields of fear conditioning research and beyond (see Supplementary Figure 3 for guidance), 2) increase rigor in reporting and analysis in the field and 3) help extinction (retention) research to resume the path for becoming "one of the best opportunities for translating neuroscience discoveries into clinical applications" (cf. 1; page 319).

Footnotes

[1]We acknowledge recent discussions suggesting the term 'threat conditioning" (46). The majority of studies included here used "fear conditioning". Hence, using a different term may lead to confusion.

[2](Ref. 26: ERI 2 and ERI 6, Ref. 27: ERI 9a and 9b, Ref. 28: ERI 15 and 16) and we only discuss CS+ based indices (excluding CS- based ERI N° 9b in (27)).

Acknowledgements

The authors would like to thank Moritz Held for the help in literature search and data organization, Habiba Schiller for help on literature search and data processing, and Eduard Focadell for sharing the data from his study. We would also like to thank all authors that provided additional clarifications on their work and the ERIs used (M. R. Milad, K.G., Martinez Gonzales, C. Hartley, B. Graham).

We also acknowledge that the idea to this manuscript is based on discussions in the context of the *European Interdisciplinary Study on Fear Conditioning and Extinction Learning as well as the return of fear* (EIFEL-ROF). We thank Anna Gerlicher, Jan Haaker and Manuel Kuhn for helpful comments on earlier versions of this article.

This work was supported by a scientific network grant from the German Research Foundation (DFG) to T.B.L. (Grant ID LO 1980/2-1) and to C.J.M. (project A09 of the SFB1280: Extinction Learning) and by a Carlos III Health Institute/FEDER grant to MAF (grant PI16/00144).

Conflicts of interest and financial disclosures

The authors declare no biomedical financial interests or potential conflicts of interest.

References

1. Anderson KC, Insel TR (2006): The promise of extinction research for the prevention and treatment of anxiety disorders. *Biol Psychiatry*. 60: 319–321.

2. Quirk GJ, Garcia R, González-Lima F (2006): Prefrontal mechanisms in extinction of conditioned fear. *Biol Psychiatry*. 60: 337–343.

3. Rauch SL, Shin LM, Phelps EA (2006): Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research--past, present, and future. *Biol Psychiatry*. 60: 376–382.

4. Sotres-Bayon F, Cain CK, LeDoux JE (2006): Brain mechanisms of fear extinction: historical perspectives on the contribution of prefrontal cortex. *Biol Psychiatry*. 60: 329–336.

5. Hermans D, Craske MG, Mineka S, Lovibond PF (2006): Extinction in human fear conditioning. *Biol Psychiatry*. 60: 361–368.

6. Myers KM, Davis M (2006): Mechanisms of fear extinction. *Mol Psychiatry*. 12: 120–150.

7. Milad MR, Quirk GJ (2012): Fear Extinction as a Model for Translational Neuroscience: Ten Years of Progress. *Annual Review of Psychology*. 63: 129–151.

8. Vervliet B, Craske MG, Hermans D (2013): Fear extinction and relapse: state of the art. *Annu Rev Clin Psychol*. 9: 215–248.

9. Dunsmoor JE, Niv Y, Daw N, Phelps EA (2015): Rethinking Extinction. *Neuron*. 88: 47–63.

10. Dymond S, Dunsmoor JE, Vervliet B, Roche B, Hermans D (2015): Fear Generalization in Humans: Systematic Review and Implications for Anxiety Disorder Research. *Behav Ther*. 46: 561–582.

11. Mataix-Cols D, Cruz LF de la, Monzani B, Rosenfield D, Andersson E, Pérez-Vigil A, *et al.* (2017): D-Cycloserine Augmentation of Exposure-Based Cognitive Behavior Therapy for Anxiety, Obsessive-Compulsive, and Posttraumatic Stress Disorders: A Systematic Review and Meta-analysis of Individual Participant Data. *JAMA Psychiatry*. 74: 501–510.

12. Raij T, Nummenmaa A, Marin MF, Porter D, Furtak S, Setsompop K, Milad MR (2018): Prefrontal Cortex Stimulation Enhances Fear Extinction Memory in Humans. *Biological psychiatry, Biological psychiatry*. 84: 129–137.

13. Lonsdorf TB, Menz MM, Andreatta M, Fullana MA, Golkar A, Haaker J, *et al.* (2017): Don't fear "fear conditioning": Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neurosci Biobehav Rev*. . doi: 10.1016/j.neubiorev.2017.02.026.

14. Bouton ME (2004): Context and behavioral processes in extinction. *Learn Mem*. 11: 485–494.

15. Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL (2007): Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol Psychiatry*. 62: 446–454.

16. Quirk GJ (2002): Memory for extinction of conditioned fear is long-lasting and persists following spontaneous recovery. *Learn Mem*. 9: 402–407.

17. Milad MR, Orr SP, Pitman RK, Rauch SL (2005): Context modulation of memory for fear extinction in humans. *Psychophysiology*. 42: 456–464.

18. Milad MR, Quinn BT, Pitman RK, Orr SP, Fischl B, Rauch SL (2005): Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *PNAS*. 102: 10706–10711.

19. Rauch SL, Milad MR, Orr SP, Quinn BT, Fischl B, Pitman RK (2005): Orbitofrontal thickness, retention of fear extinction, and extraversion. *Neuroreport*. 16: 1909–1912.

20. Forcadell E, Torrents-Rodas D, Vervliet B, Leiva D, Tortella-Feliu M, Fullana MA (2017): Does fear extinction in the laboratory predict outcomes of exposure therapy? A treatment analog study. *Int J Psychophysiol*. 121: 63–71.

21. Merz CJ, Hamacher-Dang TC, Wolf OT (2016): Immediate extinction promotes the return of fear. *Neurobiol Learn Mem*. 131: 109–116.

22. Merz CJ, Hamacher-Dang TC, Stark R, Wolf OT, Hermann A (2018): Neural Underpinnings of Cortisol Effects on Fear Extinction. *Neuropsychopharmacology*. 43: 384–392.

23. Szymanski J, O'Donohue W (1995): Fear of Spiders Questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*. 26: 31–34.

24. Spielberger CD, Gorsuch RL, Lushene RE (1983): *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

25. Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 57: 289–300.

26. Pace-Schott EF, Tracy LE, Rubin Z, Mollica AG, Ellenbogen JM, Bianchi MT, *et al.* (2014): Interactions of time of day and sleep with between-session habituation and extinction memory in young adult males. *Exp Brain Res*. 232: 1443–1458.

27. Raio CM, Brignoni-Perez E, Goldman R, Phelps EA (2014): Acute stress impairs the retrieval of extinction memory in humans. *Neurobiol Learn Mem*. 112: 212–221.

28. Pineles SL, Nillni YI, King MW, Patton SC, Bauer MR, Mostoufi SM, *et al.* (2016): Extinction retention and the menstrual cycle: Different associations for women with posttraumatic stress disorder. *J Abnorm Psychol*. 125: 349–355.

29. Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, *et al.* (2009): Neurobiological Basis of Failure to Recall Extinction Memory in Posttraumatic Stress Disorder. *Biol Psychiatry*. 66: 1075–1082.

30. Dawson ME, Schell AM, Filion DL, Berntson GG (2007): The Electrodermal System. *Handbook of Psychophysiology*, Third edition. Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/CBO9780511546396.007.

31. Baldwin SA (2017): Improving the rigor of psychophysiology research. *International Journal of Psychophysiology*, Rigor and Replication: Towards Improved Best Practices in Psychophysiological Research. 111: 5–16.

32. Ioannidis JPA (2005): Why Most Published Research Findings Are False. *PLOS Medicine*. 2: e124.

33. Simmons JP, Nelson LD, Simonsohn U (2011): False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*. 22: 1359–1366.

34. Plendl W, Wotjak CT (2010): Dissociation of within- and between-Session Extinction of Conditioned Fear. *J Neurosci*. 30: 4990–4998.

35. Prenoveau JM, Craske MG, Liao B, Ornitz EM (2013): Human fear conditioning and extinction: timing is everything…or is it? *Biol Psychol*. 92: 59–68.

36. Shumake J, Furgeson-Moreira S, Monfils MH (2014): Predictability and heritability of individual differences in fear learning. *Anim Cogn*. 17: 1207–1221.

37. Bouton ME, Westbrook RF, Corcoran KA, Maren S (2006): Contextual and temporal modulation of extinction: behavioral and biological mechanisms. *Biol Psychiatry*. 60: 352–360.

38. Bouton ME, García-Gutiérrez A, Zilski J, Moody EW (2006): Extinction in multiple contexts does not necessarily make extinction less vulnerable to relapse. *Behav Res Ther*. 44: 983–994.

39. Forcadell E, Torrents-Rodas D, Treen D, Fullana MA, Tortella-Feliu M (2017): Attentional Control and Fear Extinction in Subclinical Fear: An Exploratory Study. *Front Psychol*. 8. doi: 10.3389/fpsyg.2017.01654.

40. Straus LD, Norman SB, Risbrough VB, Acheson DT, Drummond SPA (2018): REM sleep and safety signal learning in posttraumatic stress disorder: A preliminary study in military veterans. *Neurobiol Stress*. 9: 22–28.

41. Acheson D, Feifel D, de Wilde S, McKinney R, Lohr J, Risbrough V (2013): The effect of intranasal oxytocin treatment on conditioned fear extinction and recall in a healthy human sample. *Psychopharmacology (Berl)*. 229: 199–208.

42. Acheson DT, Eyler LT, Resovsky J, Tsan E, Risbrough VB (2015): Fear extinction memory performance in a sample of stable, euthymic patients with bipolar disorder. *J Affect Disord*. 185: 230–238.

43. Morriss J, Hoare S, van Reekum CM (2018): It's time: A commentary on fear extinction in the human brain using fMRI. *Neurosci Biobehav Rev*. 94: 321–322.

44. Haaker J, Golkar A, Hermans D, Lonsdorf TB (2014): A review on human reinstatement studies: an overview and methodological challenges. *Learn Mem*. 21: 424–440.

45. Vanbrabant K, Boddez Y, Verduyn P, Mestdagh M, Hermans D, Raes F (2015): A new approach for modeling generalization gradients: a case for hierarchical models. *Front Psychol*. 6: 652.

46. LeDoux JE, Pine DS (2016): Using Neuroscience to Help Understand Fear and Anxiety: A Two-System Framework. *Am J Psychiatry*. 173: 1083–1093.

Table 1: Operationalizations and calculations of the extinction retention index (ERI) based on skin conductance responding in the literature as derived from a systematic literature search (until October 2018).

Note that some experimental protocols employed two different CS+ types during fear acquisition training, one of which was extinguished (CS+e) and one that was not (CS+u, unextinguished, sometimes also referred to as CS+ not extinguished and indicated as CS+ne). Importantly, during the retention test, both CS+ stimuli (Cs+e and CS+u) as well as the CS- are typically presented. Similarly, in studies employing a CS+ and a CS-, both stimuli are presented again during the retention test.

			Specifications used in calculation of the ERI							
Index N°	**Term used by authors**	**Formula for calculation**	**Trials used to assess retention**	**Trial type (retention)**	**Acquisition correction**	**Extinction correction**	***100**	**Division from 100 or 1**	**Recall in extinction context**	**Studies where it was used**
non-differential indices (CS+-based)										
1	% conditioned response recovered (17) Extinction retention index (18, 19)	100-[100* first CS+ of retention/Max(CS+ acquisition)]	First	CS+	Max(CS+)	no	x	x	yes	(17)[A] (18, 19)
2	Extinction retention index	100-[100*MEAN(first 2 CS+ of retention)/Max(CS+ acquisition)] OR in other experimental designs 100-[100*MEAN(first 2 CS+e of retention)/Max(CS+e acquisition)]	First 2	CS+[A] OR CS+e[A]	Max(CS+)[A] OR Max(CS+e)[A]	no	x	x	yes Except for (29)	(30, 31) (32–34)[A] (24, 35–37) (15, 29)[A]

3	Extinction retention index	100*[1-[MEAN(first 2 CS+ of retention)/(MEAN two largest CS+ during acquisition)]	First 2	CS+	2 Max(CS+)	no	x[B]	x[B]	no	(38)
4	Extinction retention index	100-[100*MEAN(first 2 CS+e of retention)/Max(acquisition)][C]	First 2	CS+e	Max(acq)	no	x	x	yes	(39)
5	Extinction recall index/recovery index	100*MEAN(first 2 CS+ trials of retention)/Max(CS+ acquisition)	First 2	CS+	Max(CS+)	no	x	no	no	(40, 41)
6	Extinction retention index	100-[100*MEAN(first 4 CS+e of retention)/Max(CS+e acquisition)] [A,E,F]	First 4	CS+e[A,E]	Max(CS+e) [A, E]	no	x[F]	x[F]	yes	(27, 42, 43)[E] (44)[A] (24, 45) (46)[F]
7	Extinction retention index	100-[100*MEAN(first 4 CS+e of retention)/Max(to a CS+ trial in acquisition)]	First 4	CS+e	Max(CS+e and CS+u)	no	x	x	yes	(47)
8	% fear recovery	100*MEAN(first 4 CS+ trials of retention)/Max(CS+ acquisition)[G]	First 4	CS+	Max(CS+)	no	x	no	no (48) yes (49)	(48, 49) [G]
9a	Retrieval index	(first CS+ during retention) – (last CS+ during extinction) [H]	First	CS+	no	last CS+[H]	no	no	no	(25)
non-differential indices (CS--based)										

9b		(first CS- during retention) – (last CS- during extinction) [H]	First	CS-	no	last CS-[H]	no	no	no	(25)
differential indices										
10	Extinction retention index	100-(100*[(MEAN first 2 CS+ of retention) – (MEAN first two CS- of retention during retention)]/Max pair[I](CS+) – (CS-) acquisition)	First 2	(CS+)-(CS-)	Max[pair[I](CS+)-(1CS-)]	no	x	x	yes	(50, 51)
11	Extinction recall index	MEAN (first 2 CS+ trials of retention) – MEAN (first 2 CS- trials of retention)	First 2	(CS+)-(CS-)	no	no	no	no	yes	(20, 52)
12	Percentage suppression (extinction) rate	100*[(MEAN CS- of retention)-(MEAN CS+ of retention)]/(MEAN CS- of retention)	All (i.e., 3)	(CS+)-(CS-)	no	no	x	no	no	(53)
13	Extinction retention index / recovery index	100-[100*MEAN((first 4 CS+ of retention) – (MEAN(first 4 CS- of retention))/(Max pair[I] (CS+) –(CS-) acquisition)]	First 4	(CS+)-(CS-)	Max[pair[I](CS+)-(1CS-)]	no	x	x	no	(54, 55)
14	Extinction recall index	MEAN((first 4 CS+u of retention) – (MEAN(first 4 CS+e of retention))	First 4	(CS+e)-(CS+u)	no	no	no	no	no[K]	(56)
15	Extinction retention score	[MEAN(first 5 CS+ trials of retention) – MEAN(first 5 CS- trials of retention)] – [MEAN(trial 2-5 CS+ of extinction) – MEAN(trial 2-5 CS- of extinction)] –	First 5	(CS+)-(CS-)	no	CS+(early extinction) – CS-(early extinction)	no	no	no	(26)

16	Extinction retention score	[MEAN(first 5 CS+ trials of retention) – MEAN(first 5 CS- trials of retention)] – [MEAN(last 5 CS+ of extinction) – MEAN(last 5 CS- of extinction)] –	First 5	(CS+)-(CS-)	no	CS+(end extinction) – CS-(end extinction)	no	no	no	(26)

Legend: differential index: based on CS+/CS- discrimination, non-differential index: based on one CS type only, CS+: conditioned stimulus +, CS-: non-conditioned stimulus, CS+e: extinguished CS+, CS+u: unextinguished CS+.

[A] The original publications referred to 'recall trial' and "maximum during acquisition" or "a CS+ trial". In these studies, "recall trial" refers to the CS+/CS+e and "maximum during acquisition" or "a CS+ trails" to the CS+/CS+e during acquisition (M.R. Milad, K.G. Martinez Gonzales, B. Graham, C. Hartley, personal communications).

[B] Note that the sequence of the terms in the formula is different from other indices

[C] "block" defined as two subsequent CS+

[D] unpublished study

[E] The original publication refers to "CS+ trial" and "maximum CS+ responding during acquisition". In this study, "CS+trial" refers to the CS+e and "maximum CS+ responding during acquisition" to the CS+e. (M.R. Milad, personal communication).

[F] The formula reported in the original publication was spelled out incorrectly (M.R. Milad, personal communication)

[G] The formulation in the publication (49) was "each subject's average SCRs during extinction recall were divided by their largest SCR to the CS+ trials during conditioning ". It was clarified by the authors that this refers to the first 4 CS+e trials during retention test (despite the retention phase having 8 trials in total) and that CS+ during conditioning referred to the CSe only (M.R. Milad & B. Graham, personal communication)

[H] The first trial of the re-extinction session was designated as a CS− to absorb the initial orienting response that commonly occurs at the start of the session, and was therefore disregarded before all day 2 analyses.' (cf. 25)

[I] "Pair" is defined as CS+ and its corresponding CS-

[K] The methods section section does not indicate contextual manipulations but refers to a previous study (15) which did employ context changes.

Supplementary Methods to

Fear extinction retention – is it what we think it is?

Tina B. Lonsdorf, Christian J. Merz & Miguel A. Fullana

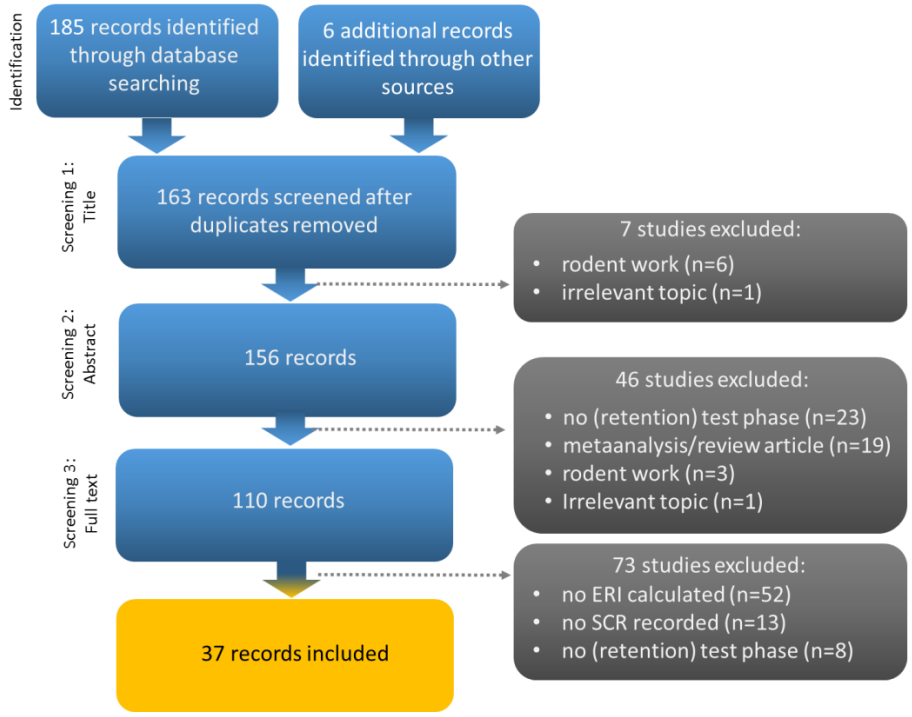
1 Supplementary Methods:

1.1 Systematic Review

A comprehensive literature search using PubMed was conducted for English-language peer-reviewed empirical studies in humans, in which an "extinction retention index" (ERI) was calculated based on skin conductance response (SCR) until 1st October, 2018. Returned articles were also manually inspected for additional studies.

The search terms for separate searches were "fear conditioning & extinction recall", "fear conditioning & extinction retention index", "extinction retention index" and "extinction recall index". We followed PRISMA guidelines (1).

The PubMed search, which was restricted to human work, yielded a total of 185 results. Six additional records were identified through other sources, of which ultimately 37 studies were included. Supplementary Figure 1 depicts a flowchart illustrating data selection and exclusion including intermediate steps.



Supplementary Figure 1: Flowchart of study selection process.

Supplementary Table 1. Overview on sample and experimental design details in the studies included in the systematic literature search.

Note that any experimental phase following retention test (for instance renewal or reinstatement) is not included here for conciseness.

Only outcome measures printed in bold were used for calculation of the extinction retention index (ERI).

Abbreviations: **acq**: acquisition training, **CS+**: conditional stimulus paired with the US, **CS+e**: CS+ extinguished, **CS+u**: CS+ unextinguished, **CS-**: CS not paired with the US, **ext**: extinction training, **fMRI**: functional magnetic resonance imaging, **FPS**: fear potentiated startle, **N°:** number, **N**: number of participants, **OCD**: obsessive-compulsive disorder, **PTSD**: post-traumatic stress disorder, **RI**: Reinforcement, **ret**: retention test, **SCR:** skin conductance responses

Index N°	Formula for calculation [A]	Ref [B]	N	sample [C]	RI rate	US type	type [D] of CS	context	immediate vs. delayed extinction	N° trials acq	N° trials ext	N° trials ret	context sequence	Outcome measure
1	100-[100* first CS+ of retention/Max(CS+ acquisition)]	(2)	30	healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(3)	14	healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(4)	14	healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
2	100-[100*MEAN(first 2 CS+ of retention)/Max(CS+ acquisition)] OR 100-[100*MEAN(first 2 CS+e of retention)/Max(CS+e acquisition)]	(5)	37 18	OCD healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(6)	14[E] 14[E]	trauma exposed healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(7)	28 18	schizophrenic healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(8)	18	healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(9)	46	healthy	100%	electric	lamp color	light rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(10)	21 21	OCD healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
		(11)	27 20	ADHD healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
		(12)	28	healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR**

		(13)	69	healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR**
		(14)	17	healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR**
		(15)	28	healthy	35%	acoustic	squares	---	delayed	23 CS+e 23 CS+u 15 CS-	30 CS+e 30 CS-	20 CS+e 20 CS+u 20 CS-	AAA	**SCR,** fMRI [F], US expectancy [F]
3	100*[1-[MEAN(first 2 CS+ of retention)/(MEAN two largest CS+ during acquisition)]	(16)	12	healthy	35%	electric	squares	---	immediate	23 CS+ 15 CS-	15 CS+ 15 CS-	17 CS+ 17 CS-	AAA	**SCR**
4	100-[100*MEAN(first 2 CS+ of retention)/Max(acquisition)]	(17)	96 [G]	social phobic	100%	electric	lamp color	light rooms	delayed	5 CS+ 5 CS-	5 CS+ 5 CS-	5 CS+ 5 CS-	ABB	**SCR**
5	100*MEAN(first 2 CS+ trials of retention)/Max(CS+ acquisition)	(18)	61	healthy	62.5%	electric	pictures of spiders	----	immediate	8 CS+ 8 CS-	7 CS+ 7 CS-	7 CS+ 7 CS-	AAA	**SCR,** US expectancy [F], valence ratings [F]
		(19)	64	healthy	62.5%	electric	male faces (neutral expression)	----	immediate	8 CS+ 8 CS-	7 CS+ 7 CS-	7 CS+ 7 CS-	AAA	**SCR**
6	100-[100*MEAN(first 4 CS+e of retention)/Max(CS+e acquisition)]	(20)	31 25	PTSD healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 8 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
		(21)	34	healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
		(22)	19 20	PTSD trauma exposed	60%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
		(23)	20	schizophrenic	60%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
		(12)	28	healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR**
		(24)	84 [H]	healthy	62.5%	electric	lamp color	light rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]

		(25)	24 20 21	PTSD trauma exposed control	62.5%	electric	lamp light color	rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
7	100-[100*MEAN(first 4 CS+e of retention)/Max(to a CS+ trial in acquisition)]	(26)	14 13	insomnia patients good sleeper	62.5%	electric	lamp light color	rooms	immediate	8 CS+e 8 CS+u n/a CS-	16 CS+e -- 16 CS-	n/a CS+e n/a CS+u n/a CS-	ABB	**SCR, subjective ratings,** fMRI [F]
8	100*MEAN(first 4 CS+ trials of retention)/Max(CS+ acquisition)[G]	(27)	83	healthy	62.5%	electric	male faces (neutral expression)	----	immediate	8 CS+ 8 CS-	7 CS+ 7 CS-	7 CS+ 7 CS-	AAA	**SCR,** US expectancy ratings [F], valence ratings [F]
		(28)	13	healthy	62.5%	electric	lamp light color	rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR**
9a	(first CS+ during retention) – (last CS+ during extinction) [H]	(29)	52	healthy	37.5%	electric	abstract fractals	----	immediate	16 CS+ 10 CS-	10 CS+ 10 CS-	10 CS+ 10 CS-	AAA	**SCR**
9b	(first CS- during retention) – (last CS- during extinction) [H]	(29)												
10	100-(100*[(MEAN first 2 CS+ of retention) – (MEAN first two CS- of retention during retention)]/Max pair(CS+) – (CS-) acquisition)	(30)	42	healthy	100%	electric	lamp light color	rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
		(31)	45	healthy	100%	electric	lamp light color	rooms	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	ABB	**SCR**
11	MEAN (first 2 CS+ trials of retention) – MEAN (first 2 CS- trials of retention)	(32)	50	spider phobics	100%	electric	lamp light color	rooms	immediate	8 CS+ 8 CS-	12 CS+ 12 CS-	6 CS+ 6 CS-	ABB	**SCR, FPS, US expectancy**
		(33)	50	spider phobics	100%	electric	lamp light color	rooms	immediate	8 CS+ 8 CS-	12 CS+ 12 CS-	6 CS+ 6 CS-	ABB	**SCR, FPS, US expectancy**
12	100*[(MEAN CS- of retention)-(MEAN CS+ of retention)]/(MEAN CS- of retention)	(34)	60	healthy	100%	electric	geometric figures	---	delayed	10 CS+1 [I] 10 CS+2 10 CS-	10 CS+1 10 CS+2 [I] 10 CS-	3 CS+1 3 CS+2 3 CS-	AAA	**SCR**
13	100-[100*MEAN((first 4 CS+ of retention) – (MEAN(first 4 CS- of retention))/(Max pair (CS+) –(CS-) acquisition)]	(35)	72	healthy	100%	acoustic	geometric figures	---	delayed	12 CS+ 12 CS-	12 CS+ 12 CS-	12 CS+ 12 CS-	AAA	**SCR**

		(36)	40	healthy	100%	acoustic	geometric figures	---	delayed	12 CS+ 12 CS-	12 CS+ 12 CS-	12 CS+ 12 CS-	AAA	**SCR**
14	MEAN((first 4 CS+u of retention) – (MEAN(first 4 CS+e of retention))	(37)	30 28	PTSD trauma exposed	62.5%	electric	lamp light color	rooms	immediate	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 16 CS-	ABB	**SCR,** fMRI [F]
15 **16**	[MEAN(first 5 CS+ trials of retention) – MEAN(first 5 CS- trials of retention)] – [MEAN(trial 2-5 CS+ of extinction) – MEAN(trial 2-5 CS- of extinction)] – [MEAN(first 5 CS+ trials of retention) – MEAN(first 5 CS- trials of retention)] – [MEAN(last 5 CS+ of extinction) – MEAN(last 5 CS- of extinction)] –	(38)	32	trauma exposed	100%	electric	shapes	--	immediate	5 CS+ 5 CS-	10 CS+ 10 CS-	5 CS+ 5 CS-	AAA	**SCR**

[A] Note that some clarifications on the formulas (e.g., definition of "pair") can be found in the legend of Table 1 in the main manuscript.

[B] Note that the reference number in this table deviates from the reference number in the main text.

[C] Note that no information on subgroups based on experimental manipulations (for instance: pharmacological challenge or stress induction) are included in the table for conciseness.

[D] Note that all CS and context types in the included studies consisted of static pictures.

[E] The sample consist of 14 pairs of monozygotic twins, whereof one twin of each pair was Vietnam veteran.

[F] Note that an ERI was not calculated for this specific outcome measure.

[F] Publication of the study plan for a clinical trial; no data are presented.

[H] Re-analysis of a pooled sample from (21) and (28).

[I] Only this CS+ was reinforced (i.e., one CS+ during acquisition, one CS+ during extinction). The extinction phase is also referred to as a second learning phase in this publication.

1.2. Re-analyses of datasets

The dataset included in the main manuscript (re-analysis of (33)), is complemented by three additional datasets (39, 40) to exclude that the results presented in the main manuscript are specific to one particular dataset and its specific experimental design choices. Please note that these additional datasets were derived from two publications which were not included in the systematic literature search as no ERI was calculated (39, 40). Supplementary Table 2 provides an overview of the main design and sample specifications of all four studies, to which the different ERI calculations were applied, while details are provided below.

Supplementary Table 2. Overview of the main design and sample specifications for the included datasets.

Abbreviations: acq: acquisition, BOLD: blood oxygenation level dependent, ext: extinction, FPS, fear-potentiated startle, ret: retention, RI: reinforcement, SCR: skin conductance response

Reference for dataset	N	immediate vs. delayed extinction	RI rate	CS types	N° trials acq	N° trials ext	N° trials ret	outcome measures
[1] Forcadell et al. (2017)	50	immediate	100%	CS+, CS-	8 CS+ 8 CS-	12 CS+ 12 CS-	6 CS+ 6 CS-	SCR, FPS, ratings
[2] Merz et al. (2018)	22	delayed	62.5%	CS+e, CS+u, CS-	8 CS+e 8 CS+u 8 CS-	16 CS+e - 16 CS-	8 CS+e 8 CS+u 8 CS-	SCR, BOLD responses, ratings
[3] Merz et al. (2016)	20	immediate	62.5%	CS+e, CS+u, CS-	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	5 CS+e 5 CS+u 5 CS-	SCR, BOLD responses, ratings
[4] Merz et al. (2016)	17	delayed	62.5%	CS+e, CS+u, CS-	8 CS+e 8 CS+u 16 CS-	16 CS+e - 16 CS-	5 CS+e 5 CS+u 5 CS-	SCR, BOLD responses, ratings

1.2.1 Dataset 1 : Re-analyses of Forcadell et al. (2017)

The data presented in the main manuscript are based on a re-analysis of Forcadell et al. (33) (note that the reference number is different in the main manuscript). This paradigm was an adaptation of that used in (2) including SCRs as the only measure of conditioned fear. In Forcadell et al (33), two other measures (US expectancies and FPS) were added. The visual contexts were photographs of two different rooms (acquisition context, CX+; extinction context, CX−) containing a lamp that switched on to one of two different colors (blue or yellow), which were the CSs (CS+ and CS−). Contexts and CSs were displayed on a computer monitor in front of the participant. On day 1, a fear acquisition training phase (in CX+) was followed by an extinction training phase (in CX−). During fear acquisition training, the CS+ co-terminated with an electric shock (US). The US was individually adjusted before the experiment (day 1) presenting shocks of gradually increasing intensity until a 'definitely annoying but not painful' shock was selected. Participants were not instructed about the CS-US contingency. During extinction training (immediately after fear acquisition training), the CS+ was not followed by the US. The CS− was never followed by the US. The extinction training phase was divided in two equal parts by a 1-minute pause (early and late extinction training). Day 2 consisted of an extinction retention phase in CX−. During day 2, the CS+ and the CS− were never followed by the US. The US was not recalibrated during day 2.

Each trial of the experiment started with presentation of the context for 10, 12 or 14s. Then, the CS was presented (i.e. the lamp switched on) for 8s, and a startle probe (50ms duration, 100dB) was delivered 7s after CS onset. Between trials, a fixation cross was shown for 1s. In one third of the trials (noise-alone trials, NA), no CS was presented, and instead the context was present for eight more seconds; the startle probe was presented at second 7 of this extra time. The inter-probe interval varied between 18, 20, and 22s. Eight trials of each type (CS+, CS−, and NA) were presented during fear acquisition training, and six trials of each were presented during each of the remaining phases (early

and late extinction training, and extinction retention; cf. Supplementary Table 1). SCR, FPS and US expectancy ratings were calculated for each trial type.

The SCR signal was sampled at a rate of 125Hz. SCR magnitudes were computed in microSiemens (μS) as the difference between the maximum SCR value and the value at response onset, occurring 1 to 7s after CS onset. Trials in which no response was detected or with a response magnitude of <0.02 μS were considered non-response trials), and trials showing interference or excessive baseline activity (1.3%) were rejected after visual inspection. To normalize the distribution of the SCR data, a square root transformation was applied.

For further details, we refer to the original publication.

1.2.2 Dataset 2: Re-analysis of Merz et al (2018)

This dataset is based on a re-analysis of Merz et al. [(39) note that the reference number is different in the main manuscript]. The visual contexts were photographs of two different rooms (acquisition context, CX+; extinction context, CX−) containing a lamp that switched on to one of three different colors (blue, red or yellow), which were the CSs (CS+e, CS+u and CS−). Contexts and CSs were presented via fMRI-ready goggles. On day 1, fear acquisition training took place in CX+, during which both CS+ (CS+e and CS+u) co-terminated with an electric shock (US) in five out of eight trials (62.5% reinforcement rate). The US was individually adjusted before the experiment presenting shocks of gradually increasing intensity until an 'unpleasant but not painful' shock was selected. Participants were not instructed about the CS-US contingency, but all of them learned the correct CS-US contingencies during fear acquisition training as evidenced by a post-acquisition questionnaire.

On day 2, participants underwent delayed extinction training (in CX−), during which the CS+e was not followed by the US. The CS− was never followed by the US and the CS+u was not presented at all. Day 3 took place one week after fear extinction training and consisted of a retention phase in CX− and a new context (note that only data of extinction retention in CX- are included in this re-analysis). During day 3, the CS+e, CS+u and the CS− were never followed by the US. The US was not recalibrated during day 2 or day 3.

Each trial of the experiment started with the presentation of a black screen between 0 and 1.875s, after that the context was shown for 3s. Then, the CS was presented (i.e. the lamp switched on) for 6s. Between trials, a black screen was shown between 9.125 and 11s (total trial duration: 20s). Eight trials of each type (CS+e, CS+u, and CS−) were presented during fear acquisition training and extinction retention, and 16 trials of CS+e and CS- were presented during extinction training (cf. Supplementary Table 1). In addition to SCRs, the BOLD signal was measured using fMRI.

The SCR signal was sampled at a rate of 5000Hz and low-pass filtered afterwards with a cutoff frequency of 10Hz. SCR magnitudes were computed in μS as the difference between the maximum SCR value and the value at response onset, occurring 1 to 6.5s after CS onset. To normalize the distribution of the SCR data, a square root transformation was applied.

For further details, we refer to the original publication.

1.2.3 Dataset 3 and 4: Re-analysis of Merz et al (2016)

This dataset is based on a re-analysis of Merz et al. (40) (note that the reference number is different in the main manuscript), in which two groups were included: a group undergoing immediate extinction and a group undergoing delayed extinction, which served as two additional datasets for this re-analysis. The visual contexts were photographs of two different rooms (acquisition context, CX+; extinction context, CX−) containing a lamp that switched on to one of three different colors (blue, red or yellow), which were the CSs (CS+e, CS+u and CS−). Contexts and CSs were displayed on a computer monitor in front of the participant. On day 1, fear acquisition training took place in CX+, during which

both CS+ (CS+e and CS+u) co-terminated with an electric shock (US) in five out of eight trials (62.5% reinforcement rate). The US was individually adjusted before the experiment presenting shocks of gradually increasing intensity until an 'unpleasant but not painful' shock was selected. Participants were not instructed about the CS-US contingency, but all of them learned the correct CS-US contingencies during fear acquisition training as evidenced by a post-acquisition questionnaire.

Participants underwent extinction training in CX−, during which the CS+e was not followed by the US. The CS− was never followed by the US and the CS+u was not presented at all. Extinction training took either place immediately after acquisition training (immediate extinction) or 24h after acquisition training (delayed extinction). One day after extinction training, both groups underwent a retention phase taking place in CX− and CX+ (note that only data of extinction retention in CX- are included in this re-analysis). During recall, the CS+e, CS+u and the CS− were never followed by the US. The US was not recalibrated during day 2 or day 3.

Each trial of the experiment started with the presentation of the context for 3s. After that, the CS was presented (i.e. the lamp switched on) for 6s. Between trials, a black screen was shown between 6 and 8s. During fear acquisition training, eight trials of CS+e and CS+u as well as 16 trials of CS- were presented. During extinction training, the CS+e and CS- were shown 16 times each. During extinction retention, each of the three CS was presented 5 times (cf. Supplementary Table 1). In addition to SCRs, the US expectancy was measured after the recall phase. As these datasets included a non-identical number of CS+e and CS- trials during fear acquisition training, the mean of two CS- trials was used to calculate index 11 and index 14.

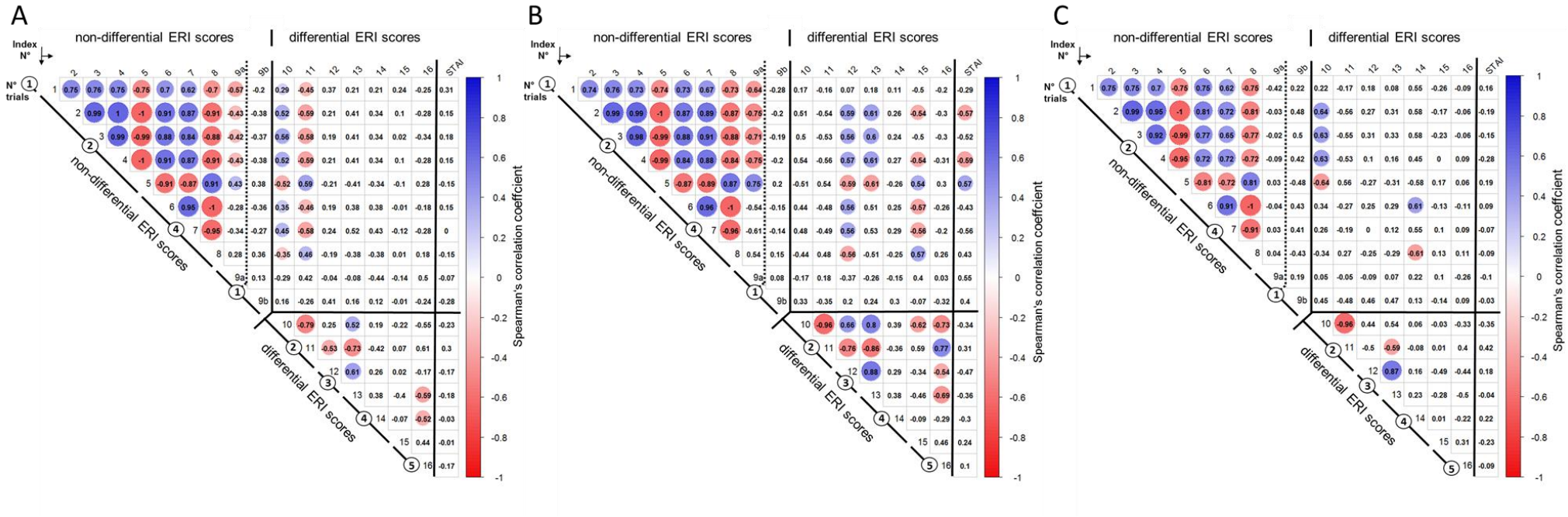
The SCR signal was sampled at a rate of 1000Hz and high-pass filtered afterwards with a cutoff frequency of 0.05Hz. SCR magnitudes were computed in µS as the difference between the maximum SCR value and the value at response onset, occurring 1 to 6.5s after CS onset. Three participants from the delayed extinction group had to be excluded due to missing data needed for the calculation of the ERI (e.g., any SCR during fear acquisition). To normalize the distribution of the SCR data, a square root transformation was applied.

For further details, we refer to the original publication.

2 Supplementary Results

The results of these additional datasets show a similar pattern as dataset 1 (see main manuscript), although number the correlations passing the significance threshold is smaller in these smaller datasets.

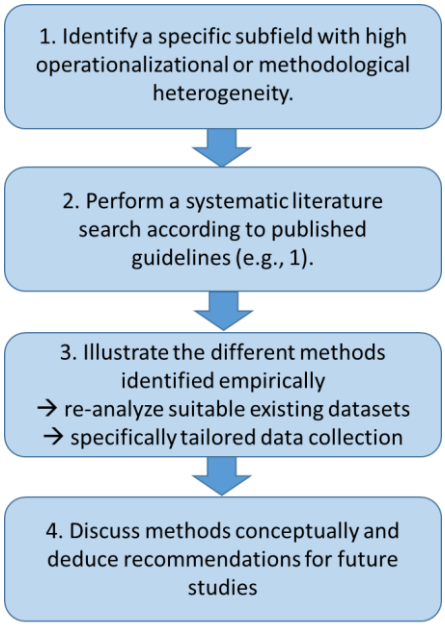
Supplementary Figure 2. Correlation matrices showing cross-ERI correlations as well as the correlation of the ERIs with the STAI trait sum score for (A) dataset 2 (N=22, delayed extinction), (B) dataset 3 (N=20, immediate extinction), and (C) dataset 4 (N=17, delayed extinction). Cells with a white background indicate non-significant correlations (p>0.05). Correction for multiple comparisons (using the BH correction, see main manuscript) was performed separately for cross-ERI correlations and correlations between ERIs and the STAI sum score in each dataset. All correlation matrices were produced as described in the main manuscript for dataset 1. Note that in dataset 3, the STAI sum score was missing from one participant.



3. Roadmap for future studies targeting methodological heterogeneity

Here, we provide a suggestion for a roadmap as a guidance for future studies addressing potentially problematic methodological heterogeneity.

Supplementary Figure 3. Suggestion for a roadmap for future studies targeting methodological heterogeneity.



Supplementary References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 6: e1000097.
2. Milad MR, Orr SP, Pitman RK, Rauch SL (2005): Context modulation of memory for fear extinction in humans. *Psychophysiology*. 42: 456–464.
3. Milad MR, Quinn BT, Pitman RK, Orr SP, Fischl B, Rauch SL (2005): Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *Proc Natl Acad Sci U S A*. 102: 10706–10711.
4. Rauch SL, Milad MR, Orr SP, Quinn BT, Fischl B, Pitman RK (2005): Orbitofrontal thickness, retention of fear extinction, and extraversion. *Neuroreport*. 16: 1909–1912.
5. McLaughlin NCR, Strong D, Abrantes A, Garnaat S, Cerny A, O'Connell C, *et al.* (2015): Extinction retention and fear renewal in a lifetime obsessive-compulsive disorder sample. *Behav Brain Res*. 280: 72–77.
6. Milad MR, Orr SP, Lasko NB, Chang Y, Rauch SL, Pitman RK (2008): Presence and Acquired Origin of Reduced Recall for Fear Extinction in PTSD: Results of a Twin Study. *J Psychiatr Res*. 42: 515–520.
7. Holt DJ, Lebron-Milad K, Milad MR, Rauch SL, Pitman RK, Orr SP, *et al.* (2009): Extinction memory is impaired in schizophrenia. *Biol Psychiatry*. 65: 455–463.
8. Zeidan MA, Lebron-Milad K, Thompson-Hollands J, Im JJY, Dougherty DD, Holt DJ, *et al.* (2012): Test-retest reliability during fear acquisition and fear extinction in humans. *CNS Neurosci Ther*. 18: 313–317.
9. Martínez KG, Castro-Couch M, Franco-Chaves JA, Ojeda-Arce B, Segura G, Milad MR, Quirk GJ (2012): Correlations between psychological tests and physiological responses during fear conditioning and renewal. *Biol Mood Anxiety Disord*. 2: 16.
10. Milad MR, Furtak SC, Greenberg JL, Keshaviah A, Im JJ, Falkenstein MJ, *et al.* (2013): Deficits in conditioned fear extinction in obsessive-compulsive disorder and neurobiological changes in the fear circuit. *JAMA Psychiatry*. 70: 608–618; quiz 554.
11. Spencer AE, Marin M-F, Milad MR, Spencer TJ, Bogucki OE, Pope AL, *et al.* (2017): Abnormal fear circuitry in Attention Deficit Hyperactivity Disorder: A controlled magnetic resonance imaging study. *Psychiatry Res Neuroimaging*. 262: 55–62.
12. Pace-Schott EF, Tracy LE, Rubin Z, Mollica AG, Ellenbogen JM, Bianchi MT, *et al.* (2014): Interactions of time of day and sleep with between-session habituation and extinction memory in young adult males. *Exp Brain Res*. 232: 1443–1458.
13. Hölzel BK, Brunsch V, Gard T, Greve DN, Koch K, Sorg C, *et al.* (2016): Mindfulness-Based Stress Reduction, Fear Conditioning, and The Uncinate Fasciculus: A Pilot Study. *Front Behav Neurosci*. 10: 124.
14. Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL (2007): Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol Psychiatry*. 62: 446–454.
15. Rabinak CA, Angstadt M, Lyons M, Mori S, Milad MR, Liberzon I, Phan KL (2014): Cannabinoid modulation of prefrontal-limbic activation during fear extinction learning and recall in humans. *Neurobiol Learn Mem*. 113: 125–134.
16. Hartley CA, Fischl B, Phelps EA (2011): Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. *Cereb Cortex N Y N 1991*. 21: 1954–1962.
17. Hofmann SG, Carpenter JK, Otto MW, Rosenfield D, Smits JAJ, Pollack MH (2015): Dose timing of D-cycloserine to augment cognitive behavioral therapy for social anxiety: Study design and rationale. *Contemp Clin Trials*. 43: 223–230.
18. Li S, Graham BM (2016): Estradiol is associated with altered cognitive and physiological responses during fear conditioning and extinction in healthy and spider phobic women. *Behav Neurosci*. 130: 614–623.
19. Milligan-Saville JS, Graham BM (2016): Mothers do it differently: reproductive experience alters fear extinction in female rats and women. *Transl Psychiatry*. 6: e928.

20. Shvil E, Sullivan GM, Schafer S, Markowitz JC, Campeas M, Wager TD, *et al.* (2014): Sex differences in extinction recall in posttraumatic stress disorder: a pilot fMRI study. *Neurobiol Learn Mem*. 113: 101–108.
21. Zeidan MA, Igoe SA, Linnman C, Vitalo A, Levine JB, Klibanski A, *et al.* (2011): Estradiol modulates medial prefrontal cortex and amygdala activity during fear extinction in women and female rats. *Biol Psychiatry*. 70: 920–927.
22. Milad MR, Pitman RK, Ellis CB, Gold AL, Shin LM, Lasko NB, *et al.* (2009): Neurobiological Basis of Failure to Recall Extinction Memory in Posttraumatic Stress Disorder. *Biol Psychiatry*. 66: 1075–1082.
23. Holt DJ, Coombs G, Zeidan MA, Goff DC, Milad MR (2012): Failure of neural responses to safety cues in schizophrenia. *Arch Gen Psychiatry*. 69: 893–903.
24. Zsido R (2014): Contributions of Estradiol and Hormonal Contraceptive Use to Sex Differences During Fear Extinction Recall. 7: 9.
25. Marin M-F, Song H, VanElzakker MB, Staples-Bradley LK, Linnman C, Pace-Schott EF, *et al.* (2016): Association of Resting Metabolism in the Fear Neural Network With Extinction Recall Activations and Clinical Measures in Trauma-Exposed Individuals. *Am J Psychiatry*. 173: 930–938.
26. Seo J, Moore KN, Gazecki S, Bottary RM, Milad MR, Song H, Pace-Schott EF (2018): Delayed fear extinction in individuals with insomnia disorder. *Sleep*. 41. doi: 10.1093/sleep/zsy095.
27. White EC, Graham BM (2016): Estradiol levels in women predict skin conductance response but not valence and expectancy ratings in conditioned fear extinction. *Neurobiol Learn Mem*. 134 Pt B: 339–348.
28. Graham BM, Milad MR (2013): Blockade of estrogen by hormonal contraceptives impairs fear extinction in female rats and women. *Biol Psychiatry*. 73: 371–378.
29. Raio CM, Brignoni-Perez E, Goldman R, Phelps EA (2014): Acute stress impairs the retrieval of extinction memory in humans. *Neurobiol Learn Mem*. 112: 212–221.
30. Milad MR, Goldstein JM, Orr SP, Wedig MM, Klibanski A, Pitman RK, Rauch SL (2006): Fear conditioning and extinction: influence of sex and menstrual cycle in healthy humans. *Behav Neurosci*. 120: 1196–1203.
31. Milad MR, Zeidan MA, Contero A, Pitman RK, Klibanski A, Rauch SL, Goldstein JM (2010): The influence of gonadal hormones on conditioned fear extinction in healthy humans. *Neuroscience*. 168: 652–658.
32. Forcadell E, Torrents-Rodas D, Treen D, Fullana MA, Tortella-Feliu M (2017): Attentional Control and Fear Extinction in Subclinical Fear: An Exploratory Study. *Front Psychol*. 8. doi: 10.3389/fpsyg.2017.01654.
33. Forcadell E, Torrents-Rodas D, Vervliet B, Leiva D, Tortella-Feliu M, Fullana MA (2017): Does fear extinction in the laboratory predict outcomes of exposure therapy? A treatment analog study. *Int J Psychophysiol Off J Int Organ Psychophysiol*. 121: 63–71.
34. Kuriyama K, Honma M, Soshi T, Fujii T, Kim Y (2011): Effect of d-cycloserine and valproic acid on the extinction of reinstated fear-conditioned responses and habituation of fear conditioning in healthy humans: a randomized controlled trial. *Psychopharmacology (Berl)*. 218: 589–597.
35. Antov MI, Stockhorst U (2014): Stress exposure prior to fear acquisition interacts with estradiol status to alter recall of fear extinction in humans. *Psychoneuroendocrinology*. 49: 106–118.
36. Antov MI, Melicherová U, Stockhorst U (2015): Cold pressor test improves fear extinction in healthy men. *Psychoneuroendocrinology*. 54: 54–59.
37. Helpman L, Marin M-F, Papini S, Zhu X, Sullivan GM, Schneier F, *et al.* (2016): Neural changes in extinction recall following prolonged exposure treatment for PTSD: A longitudinal fMRI study. *NeuroImage Clin*. 12: 715–723.
38. Pineles SL, Nillni YI, King MW, Patton SC, Bauer MR, Mostoufi SM, *et al.* (2016): Extinction retention and the menstrual cycle: Different associations for women with posttraumatic stress disorder. *J Abnorm Psychol*. 125: 349–355.

39. Merz CJ, Hamacher-Dang TC, Stark R, Wolf OT, Hermann A (2018): Neural Underpinnings of Cortisol Effects on Fear Extinction. *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol*. 43: 384–392.
40. Merz CJ, Hamacher-Dang TC, Wolf OT (2016): Immediate extinction promotes the return of fear. *Neurobiol Learn Mem*. 131: 109–116.