fortune Journals

Research Article

JOURNAL OF PSYCHIATRY AND PSYCHIATRIC DISORDERS

ISSN: 2572-519X

Reconsolidation Behavioral Updating of Human Emotional Memory: A Comprehensive Review and Unified Analysis to Identify the Causes of Replication Failures, the Role of Prediction Error, and Optimal Clinical Translation

Bruce Ecker[*]

Abstract

Memory reconsolidation (MR) has been recruited via behavioral updating procedures to achieve full annulment of a human emotional memory in over 20 studies since first reported in 2010. However, at least 14 studies have reported non-replication, the cause(s) of which have remained unclear despite extensive speculation and experimental investigation. This review examines 20 successful and 14 unsuccessful studies in detail, in an attempt to identify (a) the specific probable causes of non-replication, (b) the role of prediction error (PE), and (c) optimal clinical translation. A set of criteria is defined for principled identification of specific moments of PE and of latent cause transition in experimental procedures. Applying these criteria, a failure to induce PE is identified in every non-replication study, in most cases due to a previously overlooked element of experimental procedure. For each non-replication study, I identify a simple change of procedure that is predicted to produce PE and result in successful replication. The emerging, advanced account of PE phenomenology explains how MR produces a wide range of experimental observations, including the success of both retrieval-extinction and extinction-retrieval in driving behavioral updating. A well-defined process of unlearning and annulment of emotional memory emerges from the 20 successful studies, and the clinical translation of this potent process is illustrated by a real psychotherapy case. Lastly, this review's compendium of empirical findings is applied to evaluating previously proposed frameworks of memory reconsolidation in psychotherapy, exposing significant departures from scientific fidelity.

Keywords: Memory Reconsolidation; Behavioral Updating; Prediction Error; Latent Cause; Unlearning; Retrieval-Extinction; Extinction-Retrieval; Replication Failure; Clinical Translation; Psychotherapy; Mechanism of Change; Transformational Change.

1. Introduction

The 1997–2000 period saw the demonstration by neuroscience researchers that reactivation of a consolidated memory can induce a de-consolidated or destabilized state of the memory's encoding, or engram [1]. It was soon shown that destabilization was followed by a natural process of restabilization or reconsolidation about 6 h later, and that during that "reconsolidation window," a destabilized memory was susceptible to pharmacological disruption and annulment [2, 3, 4]. Later studies showed that behavioral procedures during

Affiliation:
Codirector, Coherence Psychology Institute, 64 Bleecker St. #253, New York, NY 10012, USA

***Corresponding author:**
Bruce Ecker, Codirector, Coherence Psychology Institute, 64 Bleecker St. #253, New York, NY 10012, USA

Citation: Ecker B. Reconsolidation Behavioral Updating of Human Emotional Memory: a Comprehensive Review and Unified Analysis to Identify the Causes of Replication Failures, the Role of Prediction Error, and Optimal Clinical Translation. Journal of Psychiatry and Psychiatric Disorders. 8 (2024): 189-265.

Received: September 24, 2024
Accepted: October 04, 2024
Published: November 01, 2024

the reconsolidation window also can produce complete unlearning and annulment of an acquired fear response in both animals [5] and humans [6].

It has been established that the mechanism of memory reconsolidation (MR) allows behavioral updating that can produce bidirectional change of memory strength, revision of specific content, or full unlearning and annulment [7, 8, 9]. It became apparent that “The reconsolidation process is crucial for the modification of existing memories and is the mechanism by which the strength and/or content of consolidated memories are updated” [8]. Given the memory-based causation of numerous clinical symptomologies of behavior, affect, cognition and somatization, MR was recognized as having potential clinical application of major importance.

Presumably each type of memory modification could have therapeutic value in some clinical circumstances. It is full annulment, however, that would have the greatest therapeutic potency, for the obvious reason that any clinical symptoms maintained entirely by a particular emotional memory would cease to occur, fully and permanently, as soon as that memory can no longer reactivate or be expressed in any way due to MR-mediated unlearning and annulment. Demonstrated therapeutic change of that kind has been referred to as “profound change” or “transformational change” [10, 11, 12] to distinguish it from the gradual, incremental, partial change that has long been the norm in the clinical field and almost the entire focus of psychotherapy outcome research [13, 14, 15].

Complete, permanent nullification of an acquired emotional response has never been observed to result from standard extinction procedures, which suppress the target response only temporarily and, in many cases, only partially [16, 17]. Several studies have determined that extinction and MR are dissociable, distinct phenomena at behavioral, neural, and molecular levels [18, 19, 20, 21]. Duvarci and Nader [19] concluded, “Reconsolidation cannot be reduced down to facilitated extinction.”

Annulment of a consolidated emotional memory has been demonstrated many times in animal [22] and human [23] studies using the two radically different types of experimental MR procedure noted above, namely endogenous versus exogenous:

- The endogenous, behavioral disruption of memory contents through unlearning rewrites and replaces target memory encoding and relies on completion of the natural reconsolidation process to establish the updated memory [24, 25, 26, 27, 28].
- The exogenous, pharmacological disruption of the reconsolidation process itself, through a blockade of the de novo protein synthesis it requires, prevents reconsolidation of the target memory [29, 30, 31, 32].

The focus of the present review is the endogenous, behavioral process, because in several respects that approach appears advantageous for clinical application:

- As Soeter and Kindt [33] stated, “Obviously, a behavioral procedure will be preferred over pharmacological manipulations provided that similar effects can be obtained.”
- The behavioral process has been shown to eliminate threat memory from both the subcortical (amygdala) emotional implicit memory system and also the expectancy memory of the neocortical declarative memory system [34, 35, 36, 37, 38], whereas the latter memory system is not cleared of threat memory pharmacologically [29, 39].
- Behavioral annulment of memory via MR, no less than pharmacological annulment, acts on the memory engram, i.e., the encoding of the target memory, as distinct from disrupting memory retrieval circuits, according to (a) the “strong evidence” derived from a review of studies of brain mechanisms of long-term memory [40] and (b) neurochemical evidence that the encoding of the target learning is reconstituted by behavioral updating [41, 42]. Those findings support use of the term “erasure” in referring to MR-induced annulment of emotional memory [43, 41, 40, 30, 44]. Importantly, episodic memory of past events is not erased by annulment of a target emotional response [29, 30]. It is not yet clear whether any of a target emotional memory’s engram survives behavioral annulment [45]. Therefore, to avoid possible misunderstanding, the present article uses the term annulment rather than erasure in order to focus only on the functional elimination of all expression of a target memory, without addressing or implying anything regarding what such annulment means about the final condition of the underlying engram.
- Early but extensive clinical applications of the behavioral updating process (described in Section 6 below), including numerous cases documented in detail in peer-reviewed articles [46, 47, 11, 48], indicate both its effectiveness and its applicability for a much wider range of clinical symptoms than has so far been documented for the pharmacological approach.
- The behavioral approach to annulment is available for use by all psychotherapists, counselors, and coaches, in all categories of licensure, whereas only a small, elite minority of clinicians can use the pharmacological approach.

For those reasons, the endogenous process of MR behavioral updating and annulment of human emotional memory is the focus of this review, with the aim of advancing the clinical application of that process. This review therefore examines all laboratory studies of human behavioral updating of emotional memory through September 2021 that have either achieved annulment of emotional memory (20 studies, listed in Table 1 and reviewed in Appendix A) or failed to replicate the effect (14 studies, listed in Table 2 and reviewed in Appendix B). The aim is to identify the probable cause(s) of non-replication and, on that basis, to identify how clinical translation might optimally be designed. For that purpose, only human studies are considered here, as in some previous reviews [29, 49, 50].

Table 1: Human behavioral updating studies that nullified an acquired emotional response, and their coverage by review articles. A detailed reviewed of each study is in the subsection of Appendix A indicated in the first column.

Study	Discuss PE?	Procedural variations and other distinctive features	This article	2019 Zuccolo, Hunziker	2018 Monfils, Holmes	2018 Elsey et al.	2018 Meir Drexler & Wolf	2017 Lee et al.	2017 Beckers & Kindt	2017 Treanor et al.	2016a Clem & Schiller	2016 Kredlow et al.
[A.1] 2010 Schiller et al.	X	First behavioral annulment of human threat memory; 38% shock US pairing; SCR; CS-reactivation brings CS-specific memory annulment; effect maintained at 1 yr	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[A.2] 2012 Oyarzún et al.	✓	Replication of Schiller et al. (2010) but with auditory aversive US	✓	✓	✓	✓	X	✓	✓	X	✓	✓
[A.3] 2012 Agren et al.	X	CS images within a visual context; 100% shock US pairing; SCR and fMRI; effect maintained at 18 months	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[A.4] 2013 Schiller et al.	X	Replicates Schiller et al. (2010) experiment 2 plus fMRI brain imaging showing that reconsolidation and extinction engage different brain regions.	✓	✓	✓	✓	✓	✓	X	X	✓	✓
[A.5] 2014 Liu et al.	X	38% shock US pairing; reactivation by weakened US updates all linked CSs; US-specific memory effects; 14-day-old memory; effects maintained at 6 months	✓	✓	✓	✓	X	✓	X	X	X	X
[A.6] 2014 Steinfurth et al.	X	Replicates Schiller et al. (2010) to test 6-day-old fear memory.	✓	✓	✓	✓	✓	✓	✓	✓	X	X
[A.7] 2014 Pine et al.	✓	Affective preference memory formed by subliminal instrumental conditioning; updating via change of contingency, not extinction.	✓	X	X	X	X	X	X	X	X	X
[A.8] 2015 Johnson & Casey	X	50% US pairing; compound aversive US: auditory + image; CS images within a visual context; compares effect in adolescents and adults	✓	✓	✓	✓	X	✓	X	X	X	X
[A.9] 2016 Asthana et al.	✓	80% aversive auditory US pairing; US-reactivation; full effect for BDNF met allele carriers, reduced effect for non-met allele.	✓	✓	X	✓	X	✓	X	X	n/a	n/a

[A.10] 2017 Golkar et al.	X	Fear-relevant CSs; 80% shock US pairing; vicarious extinction; FPS & SCR	✓	✓	✓	X	X	✓	n/a	n/a	n/a	n/a
[A.11] 2017 Thompson & Lipp	✓	Fear-relevant CSs; 100% shock US pairing; reactivation by weakened US updates all linked CSs; SCR; affective ratings of CSs and US; detailed PE analysis	✓	✓	X	✓	X	✓	n/a	n/a	n/a	n/a
[A.12] 2017 Li et al.	✓	Compound, 3-part CSs with 60% shock US pairing; use of 1/3, 2/3, 3/3, and no reminder; SCR	✓	X	X	X	X	n/a	n/a	n/a	n/a	n/a
[A.13] 2018 Hu et al.	✓	Replicates Schiller et al. (2010) experiment 1 for range of reminder durations, finding updating with short but not long reminders; SCR; roles of PE and latent causes discussed	✓	X	X	n/a	n/a	n/a	n/a	n/a	n/a	n/a
[A.14] 2018 Chen et al.	✓	50% shock US pairing; uses design of Sevenster et al. (2013, 2014) with compound CSs to test reactivation with and without PE, confirming PE requirement; SCR	✓	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
[A.15] 2019 Grégoire & Greening	X	Replicates Schiller et al. (2010) but uses internal mental image of unreinforced CS+ for reactivation-reminder; SCR	✓	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
[A.16] 2019 Junjiao et al.	✓	50% shock US pairing; uses design of Sevenster et al. (2013, 2014) with compound CSs to test reactivation novelty with and without PE, confirming PE requirement; SCR; fMRI imaging shows reconsolidation and extinction engage different brain regions and connectivity.	✓	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
[A.17] 2020 Kitamura et al.	✓	Fear response is annulled at both 40% & 80% shock US pairing in replication of experiment 2 of Schiller et al. (2010); CS-reactivation brings CS-specific memory annulment; SCR; US expectancy rating after reactivation	✓	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

[A.18] 2021a Chen et al.	✓	100% shock US pairing; each CS+ paired with two shocks; FPS & SCR; fear memories of different strengths were reactivated with different degrees of mismatch; stronger fear needs greater reactivation mismatch to create PE that induces destabilization.	✓	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
[A.19] 2021c Chen et al.	✓	50% shock US pairing; SCR; retrieval-extinction eliminates spontaneous recovery for both genders but eliminates reinstatement only for female subjects, accounting for some of the inconsistency of results across studies.	✓	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
[A.20] 2021b Chen et al.	✓	67% shock US pairing; both retrieval-extinction and extinction-retrieval; fear response tested 3 h, 12 h, and 12 h with sleep after memory manipulation, and at 24 h; SCR; US expectancy rating after short-term fear test at end of Day 2	✓	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

CS: conditioned stimulus; FPS: fear-potentiated startle measured as index of fear; PE: prediction error; SCR: skin conductance response measured as index of fear; US: unconditioned stimulus

Table 2: Human behavioral updating studies that failed to nullify an emotional learning with human subjects. A detailed reviewed of each study is in the subsection of Appendix B indicated in the first column.

Study	Discuss PE?	Procedural variations and other distinctive features	Probable causes of absence of PE and other adverse effects (explained in reviews in Appendix B)
[B.1] 2011 Soeter & Kindt	X	Fear-relevant CSs; 80% shock US pairing; FPS; SCR; US expectancy ratings	Subjects continually focused on subjective rating of their distress in response to every CS presentation, interfering with PE and unlearning.
[B.2] 2012 Golkar et al.	X	Fear-relevant and fear-irrelevant CSs; 50% shock US pairing; FPS; SCR	10-min post-reactivation delay defined as video “break” induced change of latent cause, preventing PE experience.
[B.3] 2013 Kindt & Soeter	✓	Fear-relevant CSs; 75% shock US pairing; FPS; SCR; US expectancy ratings	Subjects continually focused on subjective rating of their US expectancy in response to every CS presentation, interfering with PE and unlearning. 10-min post-reactivation delay defined as magazine “break” with electronic equipment disconnected induced change of latent cause, preventing PE experience.
[B.4] 2014 Meir Drexler et al.	X	Contextual fear memory associated with fear-relevant CSs; 75% shock US pairing; SCR; US expectancy ratings	Long-duration reactivation induced change of latent cause and extinction learning instead of PE and updating.
[B.5] 2014 Warren et al.	X	100% throat airblast US pairing; FPS; with and without US expectancy ratings	Reactivation mismatch was too weak to create PE experience.
[B.6] 2016 Fricchione et al.	✓	Fear-relevant CSs; fear-prone subjects; 62.5% shock US pairing; SCR	10-min post-reactivation wait (no video or magazines) similar to 5-min wait during acquisition procedure, so no mismatch and no PE
[B.7] 2016 Klucken et al.	X	50% shock US pairing; SCR; fMRI brain imaging	10-min post-reactivation delay defined as video “break” induced change of latent cause, preventing PE experience.

[B.8] 2017 Kroes et al.	X	50% shock US pairing with different images in same category, creating generalized threat memory; return of fear tests immediately and after 24 h; tests of episodic memory of threat learning; SCR	10-min post-reactivation delay defined as video "break" with electronic equipment disconnected induced change of latent cause, preventing PE experience.
[B.9] 2018 Fernandez-Rey et al.	✓	38% pairing with US white noise burst; post-reactivation delays of 10 and 20 min before extinction; 48 h from extinction to test; SCR	Updating effect dominated by generalization of fear to the neutral cue; 10-min video defined as "rest period" induced change of latent cause, preventing PE experience.
[B.10] 2018 Kredlow et al.	X	60% US pairing; US=shock or shock+scream; 1 or 3 days of acquisition; healthy or anxious subjects; SCR	Updating effect dominated by generalization of fear to the neutral cue; fear amplified by subjects' high level of contingency awareness; possibility of 10-min video defined as "break" inducing change of latent cause, preventing PE experience.
[B.11] 2018 Zuccolo & Hunziker	X	Attempted exact replication of Schiller et al. (2010), experiment 2; SCR	10-min post-reactivation delay defined as magazine "break" with electronic equipment disconnected induced change of latent cause, preventing PE experience.
[B.12] 2020a Chalkia et al.	X	Attempted exact replication of Schiller et al. (2010); SCR	10-min post-reactivation delay defined as TV show "break" induced change of latent cause, preventing PE experience
[B.13] 2020 Houtekamer et al.	✓	60% shock US pairing with contextual fear memory created with virtual reality goggles; FPS	10-min post-reactivation delay defined as video "break" induced change of latent cause, preventing PE experience.
[B.14] 2020 Zimmermann & Bach	✓	Close replication of Schiller et al. (2010); 50% shock US pairing; SCR; pupillary response	10-min post-reactivation delay defined as video "break" induced change of latent cause, preventing PE experience.

CS: conditioned stimulus; FPS: fear-potentiated startle measured as index of fear; PE: prediction error; SCR: skin conductance response measured as index of fear; US: unconditioned stimulus

Failure to show the updating and annulment effect with both human and animal subjects has been reported in numerous studies of the behavioral (as well as pharmacological) MR approach [51, 52, 53, 54, 31]. That inconsistency has generated an intense focus of research on identifying the ranges and the interplay of various experimental variables, i.e., the boundary conditions, that allow, or disallow, MR and annulment to occur [55, 56, 8, 57, 58, 28]. A wide range of procedural variables and mnemonic effects have been discussed as possible causes of the non-replications, in both the review articles as well as in articles reporting individual studies. To date, these many factors have remained speculative and the complex dilemma of non-replication has remained somewhat mystifying. "Replication failures and alternative explanations challenge clinical translation. These obstacles can be viewed constructively by highlighting the need to demarcate the source of failure: lack of destabilization or insufficient updating at restabilization." [26] The current review demarcates the source of failure as being a lack of destabilization due to failure to create the required prediction error (PE) experience. This is shown by closely examining the moment-to-moment procedure in each of the studies in Tables 1 and 2.

Section 2 below defines the framework used for examining and evaluating studies of human behavioral updating, and Section 3 defines criteria for detecting the occurrence of PE in experimental procedures. Sections 4 and 5 present the research review and its provisional conclusions consisting of (a) a unified explanation of all studies in Tables 1 and 2, both annulment successes and annulment failures, in terms of the presence or absence of a PE experience, and (b) identification of an invariant set of distinct experiences that was created in all successful annulment studies (Table 1), even though their concrete procedures varied widely. Sections 6 and 7 discuss and illustrate the implied methodology for clinical translation, defined as replication of that set of experiences rather than any particular procedure for creating them. Section 8 uses the findings of this research review for a scientific fidelity check-up on several proposed frameworks of the therapeutic application of MR. Section 9 lists the main conclusions emerging from this review.

This is not a formal systematic review in that it is not concerned with effect sizes or statistical analyses. Rather, it has a phenomenological, qualitative approach, consisting of a systematic, fine-grained comparative scrutiny of the concrete procedures of relevant studies and the subjective experiences induced in experimental subjects by those procedures. The latter aspect is a new, important dimension of analysis of this body of research.

The following information, normally given for any formal systematic review, is nevertheless relevant to provide here. The primary search strategy for identifying studies for inclusion in Tables 1 and 2 was to comb through the text and reference list of the behavioral updating review articles since 2016 that most thoroughly and directly covered studies of human MR behavioral updating of emotional memory, indicated in the top row of Table 1, and also several MR review articles: [55, 56, 8, 58].

In addition, the PsycINFO database was searched on 2021-09-01 using the following search terms: (human) NOT ((((((((((((propranolol) OR alcohol) OR cocaine) OR adrenergic) OR noradrenergic) OR antagonist) pharmacolog*) OR transcranial) OR declarative) OR animal) OR rats) OR mice) AND ((((((((((((behavioral updat*) OR behavioral disruption) OR behavioral) OR reconsolidation updat*) OR updat*) OR retrieval-extinction) OR retrieval extinction) OR post-reactivation extinction) OR reminder-extinction) OR extinction during reconsolidation) OR extinction training during reconsolidation) OR extinction learning during reconsolidation) AND (((((((preventing the return of fear) OR prevent return of fear) OR return of fear) OR recovery of fear) OR fear memor*) OR disrupt*) OR disrupting reconsolidation) AND ((((((((reactivat*) OR reminder) OR destabiliz*) OR destablis*) OR memory reconsolidation) OR reconsolidation) OR retrieval) OR boundary condition*). That search was limited to human studies and retrieved 310 titles, resulting in the addition of a few of the most recent studies to Tables 1 and 2.

Laboratory human studies have been included in Table 1 only if the data unambiguously show full annulment of an emotional memory, i.e., return-of-response test levels that are statistically indistinguishable from zero, whether measured as a differential response relative to the response to a neutral cue or context, or as the absolute response relative to the extinguished level immediately after the behavioral updating procedure, at least one day after the memory manipulation, in a test of spontaneous recovery (re-extinction, i.e., exposure to the conditioned stimulus) and/or reinstatement (exposure to the unconditioned stimulus followed by exposure to the conditioned stimulus). Also, the memory manipulation must have been performed at least one day after memory acquisition.

Studies excluded from Table 1 are those in which (a) fear memory acquisition used highly complex stimuli that allowed for uncontrolled effects of individual differences (specifically [59], in which fear memory was created using “a distressing film paradigm”), or (b) the data suggest full annulment but are rendered ambiguous or doubtful due to either inadequate statistics, results that are too dependent on a particular or non-standard analysis of data, or loss of return-of-response data due to mishap [34, 60, 61, 62, 63, 64].

2. A framework for examining empirical studies of human behavioral updating

This section further defines how the studies listed in Tables 1 and 2 are examined in the following sections. Two subsections follow. One addresses the fundamental matter of determining whether and when the MR mechanism has been engaged and operative in a particular human study. The other proposes to analyze studies not only in terms of their externally definable procedures, or objective process, but also in terms of their internally created experiences, or subjective process.

2.1 Determining whether an experimental or clinical procedure has induced MR

Animal studies allow the most certain evidence that destabilization and reconsolidation have occurred, through the use of pharmacological and brain biochemical assay techniques [65] that detect cellular and molecular signatures of the distinct processes of destabilization, updating, and reconsolidation [66]. In human studies, brain assay is not possible and the only feasible (because non-toxic) pharmacological agent, propranolol, has been inconsistent in its effects [23]. Consequently, researchers’ best evidence of human MR is behavioral, namely, the target memory’s characteristic behavioral, emotional, and physiological expressions are changed significantly and non-reversibly by the procedure. Stable, enduring change can be due only to a fundamental change in the encoding ensemble that stores and expresses the target memory, not to the formation of a separate, competing learning [24], and only the MR process can reconstitute that encoding ensemble, according to all current scientific knowledge.

The most unambiguous and decisive form of such stable, enduring change due to MR is the full annulment of all expression of the target memory, persisting with no use of preventative procedures or conditions and tested by cues and contexts that previously evoked memory reactivation and expression. Even that is not rigorous proof of MR, it is regarded by researchers as strong circumstantial evidence [23, 24]. The lasting disappearance of all markers of target memory reactivation and expression is the most reliable verification of recruitment of MR until such time as researchers develop a method of direct detection of human engram nullification that is safe and practical. Some progress in that direction was recently reported in a human study [67] that used two non-invasive physiological markers: event-related potentials in the brain measured via electroencephalography and pupillary responses measured via eyetracking. Those data in combination distinguished between reactivation of a memory that had never been reconsolidated, a memory that previously had been reconsolidated without updating, and a memory that previously had been reconsolidated with updating. That remarkable result lends hope for decisive detection of MR occurring in human behavioral studies.

The same logic as delineated above applies to clinical application: If a psychotherapy recipient’s specific long-term emotional memory contents and their symptomatic expressions suddenly cease to be re-evoked by the cues and contexts that consistently had been evoking them, and this disappearance persists long-term with no further preventative measures, the only tenable scientific explanation is emotional memory annulment by the MR mechanism. Numerous clinical case examples of that kind have been documented in detail [47, 68, 11, 48], showing that the cessation of emotional memory reactivation and expression is reported following the same set of experiences that Section 4 reports is present in each of the studies in Table 1. One such clinical case example is presented below in Section 6.2.

Thus, clinical utilization of MR behavioral annulment of emotional memory can have the same type and degree of verification of MR induction as in laboratory empirical MR annulment studies with human subjects.

2.2 Mapping both procedures and experiences maximizes what MR studies reveal

This article proposes that our understanding of human MR studies and their clinical translation is greatly advanced by applying the distinction, previously introduced [47], between the concrete procedures used in any given study and the subjective experiences created in subjects by that procedure. The lens of experiences, used in parallel with the lens of procedures, is found to shed new light on the interpretation of experimental results and on the implications for clinical translation. The lens of procedures identifies the externally apparent, objective features of how the experimentalists acted upon the subjects. The lens of experiences enables a phenomenological account of the internal, subjective effects produced by the procedures. Using the two lenses together results in significantly more understanding from a human behavioral updating study than through using either one alone.

When the authors of laboratory MR studies address clinical translation [29, 69, 56, 70, 71, 26, 72, 73], as a rule the discussion assumes that clinical translation requires finding how to implement in psychotherapy sessions the same procedures that were used successfully in laboratory studies. For example, Thompson and Lipp [74], in discussing their successful behavioral annulment study that used the unconditioned stimulus (US) rather than the conditioned stimulus (CS) to reactivate the target memory, state (p. 8), “an obvious challenge for clinical applications of the US-reactivation procedure is the identification of the US that contributed to the development of cue-dependent associations, as well as the adaptation of the US to resemble the reactivation procedure used in the present study.”

As can be seen in the final clause of the foregoing quote, laboratory researchers tend to impose a concrete procedure replication requirement on the clinical translation project. Imposing that requirement generates serious obstacles to clinical translation of MR research findings because the clinical situation’s inherent complexities, unknowns, and uncontrollability make replication of specific laboratory procedures usually impractical and often impossible. The various researchers who have considered clinical application, referenced at the start of the previous paragraph, have identified several specific obstacles to clinical translation that exist only within that paradigm of the concrete procedure replication requirement, but do not exist within the alternative proposed paradigm [13, 47], which could be termed the experiential process replication requirement.

The latter approach guides clinical translation to consist of replicating the series of subjective experiences had by subjects in the successful laboratory annulment studies and disregards the particular concrete procedures used for inducing those experiences in any given study. Section 4 maps out the specifics of that methodology, as extracted from laboratory MR studies, and Section 6 maps out the same methodology for use in psychotherapy. The soundness of that approach derives from the fact, stated and illustrated previously [47] but fully documented for the first time in the present article, that across all successful human behavioral annulment studies, concrete procedures varied widely but the same set of three subjective experiences was produced by the procedure in every case, as shown in the individual study reviews in Appendix A, which are summarized in Section 4.

The distinction between procedures and experiences will be utilized again in later sections to illuminate (a) the relationship between memory mismatch, understood as an objective feature of procedure, and prediction error, understood as a subjective internal experience, and (b) the reported observations that the retrieval-extinction protocol of behavioral updating has remained effective when reversed and carried out as an extinction-retrieval protocol [75, 76, 77].

3. Memory mismatch, prediction error, and latent cause in MR studies

3.1 An uphill climb for the prediction error requirement

It was recognized early on that the MR process begins when memory reactivation causes the memory to become “labile” and “destabilized” [2, 78], allowing its pharmacological disruption and nullification. The stable, consolidated memory becomes “deconsolidated” [79, 80, 81, 77]. The observation that destabilization occurs only if reactivation conditions have a mismatch or discrepancy with what is known and expected according to the target memory was first reported in 2004 by Pedreira, Pérez-Cuesta, and Maldonado [82]. Since then, more than 30 laboratory studies have been designed specifically to test the hypothesis that memory destabilization occurs only if reactivation is accompanied by such mismatch, creating a prediction error (PE) experience. (See https://bit.ly/2b8IbJH for a growing, chronological list of those confirming studies, their designs, the type of memory tested, and the species tested. To the author’s best knowledge, that list is the most comprehensive compendium of PE test studies to date.) The PE requirement has been confirmed by all of those studies, which have tested a wide range of types of memory and have used a wide range of different concrete procedures to create

PE experiences for testing that hypothesis. Some of those studies have concluded explicitly that memory reactivation alone does not induce destabilization [83, 84]. Three of the most recent additions to that list [110, 180, 117] have for the first time tested the necessity of PE in the behavioral updating of human emotional memory, again confirming the PE requirement (as reviewed in Sections A.14, A.16, and A.18). Another recent human study created a PE experience entirely as an internal mental image of the conditioned stimulus appearing without the unconditioned stimulus (one day after classical conditioning via 100% reinforcement), with subsequent updating and annulment regarded as confirmation that PE had occurred in that imaginal manner [85].

The importance of PE for MR and the extensiveness of the MR-related PE research is indicated by the fact that several review articles have focused specifically on PE in MR [86, 87, 88, 70, 89, 90]. Other review articles have made strong statements in recognition of the PE requirement. In reviewing research on reconsolidation of emotional learnings in humans, Agren [91] stated, “it would appear that prediction error is vital for a reactivation of memory to trigger a reconsolidation process.” Delorenzi et al. [92], in a review of research on the relationship between memory retrieval and memory expression, commented (p. 309), “strong evidence supports the view that reconsolidation depends on detecting mismatches between actual and expected experiences.” Lee [78] wrote, “reconsolidation is triggered by a violation of expectation based upon prior learning, whether such a violation is qualitative (the outcome not occurring at all) or quantitative (the magnitude of the outcome not being fully predicted).”

The requirement for PE is regarded by MR researchers as a boundary condition, the term used for any parameter or condition in laboratory studies that has a range outside of which memory destabilization does not occur. Target memory strength, target memory age, cue specificity, and the structure or duration of target memory reactivation by a cue or context have also been shown in various studies to function as boundary conditions [93, 41, 94, 95, 96, 97, 98, 99, 82, 4, 100, 101, 102], and several review articles have focused specifically on boundary conditions, as noted above.

Instances of those various parameters preventing destabilization, such as [101], could be explained, on close analysis, as alterations of the threshold for generating PE [46]. For example, the degree of procedural mismatch required for creating a PE experience is different for target memories of different ages. Thus, PE can be understood as being senior to the other boundary conditions in the sense that, by adjusting reactivation or post-reactivation conditions so that a PE experience is created and destabilization occurs, all other boundary conditions have been taken into account implicitly.

However, as indicated in Tables 1 and 2, a minority of the studies’ authors have included discussion of PE. Of the 34 studies listed in the two tables, only 17 mention PE at all, and only half of those 17 provide a detailed consideration of PE. In a review article, Beckers and Kindt [29] commented, “The notion that memory destabilization will occur only when there is an expectancy violation or an experienced mismatch at the time of memory retrieval...has not been taken into account in the majority of clinical trials that have been conducted so far.” Ever since it was first reported [82] that a memory mismatch was necessary for destabilization and reconsolidation to occur, and even as confirming studies steadily accumulated year by year (as shown in the chronological list of those studies at https://bit.ly/2b8IbJH), recognition of the PE requirement by reconsolidation researchers has been quite uneven. It has been strongly recognized by many, as noted above, but has received no recognition or mention by others, such as [103, 9, 6, 38, 104], and recently has even been disputed by others [105] (to which a rebuttal is given below in Section 3.2.4).

A previous analysis of some replication failures found that an absence of PE in the procedure was responsible [46]. For example, rats’ prior incentive learning (an appetitive lever-pressing memory) was reactivated with the expected reinforcement (sugar reward), after which the memory was found not to be nullified by pharmacological protein synthesis blockade (106). The authors concluded that the type of memory under study was not subject to the MR process, when the clear absence of a PE experience was a far more probable cause of memory persistence, and subsequently this type of memory was indeed nullified by post-reactivation protein synthesis blockade [107]. Similarly, a human study [108] failed to replicate the use of propranolol to abolish associative conditioning, which the authors attributed, correctly, to reactivation without creating a PE experience.

The present review extends that type of analysis in a systematic manner to all of the human behavioral updating replication failures that are listed in Table 2, reviewed in Appendix B, and discussed in Section 5.

3.2 A framework for PE evaluation in experimental procedures

Reviewing each of the studies in Tables 1 and 2 involves evaluating and comparing several categories of information, one of which is the presence or absence of a PE experience created by the procedure. As noted above, the role of PE in MR does not have a conceptual consensus among researchers, with no mention of PE or memory mismatch in half of the relevant human studies, and clear criteria have not been developed for identifying the presence or absence of PE in MR studies. Therefore, addressing PE in this review requires defining a PE evaluation framework based on the preponderance of empirical data to date and applying it uniformly to all studies.

The PE evaluation framework defined below is a principled but provisional construction that may need revision as research continues to illuminate destabilization phenomenology. It is applied in this review as an exploration of the possibility that a set of empirically based criteria for PE could identify the presence or absence of PE experiences in the studies in Tables 1 and 2, could illuminate the role of PE in determining whether a study succeeds or fails to produce updating and nullification of emotional memory, and could guide clinical application. The following trial criteria constitute the PE evaluation framework used for that exploration in this review:

- PE is necessary for destabilization and updating
- PE is relative to the entirety of target memory contents
- PE is defined to be a subjective experience that may or may not occur in response to a particular mismatch defined as a feature of the external procedure.
- Mismatches that are sub-salient (very weak) and super-salient (very strong) do not create a PE experience, graphically forming an inverted-U-shaped curve of the probability of PE and destabilization versus mismatch strength.
- PE specificity can differentially destabilize only a specific component of the target memory

The first four of those criteria are explained below:

3.2.1 PE is necessary for destabilization and updating

The preponderance of the empirical findings, described above, supports the assumption that creation of a PE experience is strictly necessary for destabilization and updating of a target memory. Yet no particular concrete procedure is required for creating a PE experience, so this criterion embodies the experiential process replication requirement, rather than the concrete procedure replication requirement, for both laboratory research and clinical application.

By this assumed criterion, if a study's data unambiguously indicate that memory updating has occurred (i.e., significant, stable change of memory expression is observed), it is valid to infer that the procedure created a PE experience prior to the learning experience(s) that drove memory updating, without making the error of circular reasoning. That PE inference principle is used in some of the reviews in Appendices A and B. It allows a reverse-engineering approach when necessary for analyzing cause and effect in experimental MR studies. In other words, regarding studies that did achieve behavioral updating, with this criterion we do not have to wonder whether a PE experience occurred, rather we know it did occur and we look for precisely where it occurred in the procedure. In studies that failed to achieve updating, we must make sense of why the specific elements of the reactivation and post-reactivation procedure did not create a PE experience.

3.2.2 PE is relative to the entirety of target memory contents

Numerous studies of PE in relation to MR have used systematic parameter variations of the acquisition training and/or reactivation conditions and in that way have shown that the entire content of a target learning, including all details of the percepts, sequences, intensities, subjective uncertainties, time durations, spatial positions, latent causes (i.e., causations implicitly construed) in the acquisition training, and environmental context, determines whether particular reactivation conditions do, or do not, create a PE experience [109, 110, 111, 112, 113, 114, 20, 115, 116, 117]. As summarized by Junjiao et al. [118], "The prediction error is defined by the interaction between the information in the reminder session and the learning history." Fernández et al. [8] stated, "the boundary conditions of the reconsolidation process [i.e., the ranges of conditions that produce destabilization] are not fixed and vary as a consequence of the interaction between memory features and reminder characteristics." Bos et al. [108], discussing their non-replication of pharmacological erasure in a human study, stated (p. 6), "The experience of a prediction error upon reactivation critically depends on the interaction between the original learning of the fear association and the memory retrieval." That principle has been recognized in most review articles on PE in MR [119, 87, 88, 70, 89, 90] and has been termed the mismatch relativity principle [46, 47], with examples of its use and value for analyzing experimental studies.

Here that principle is further developed and applied as follows in the reviews in Appendices A and B of the studies listed in Tables 1 and 2, respectively: For each study, the detailed content of the target memory is itemized as a compendium of all features of the experimental acquisition conditions that subjects perceived (as enumerated in the previous paragraph). That complete account of target memory components contains more items, as a rule, than the researchers conceptualized as constituting the intended target memory. It could be argued that the memory of that entire pattern, as a mental model of how the world behaves, is semantic memory, not episodic memory. That complete set of target memory components is then compared to all features of the reactivation condition perceived by each subject, and every such procedural mismatch is identified and noted. Each identified mismatch is then evaluated for whether it did induce a PE experience or did not induce a PE experience because its salience was either too weak or too strong to do so (as explained below).

That analysis is an exploration of whether the trial criteria of PE phenomenology being applied can consistently account for the totality of empirical findings. The result, summarized in Sections 4 and 5, is that these criteria identify a PE-inducing mismatch in all of the successful annulment studies (Table 1) and an absence of any PE-inducing mismatch in all of the unsuccessful annulment studies (Table 2).

The PE evaluation of a particular mismatch must take into account both the possible presence of other reactivation mismatches and whether or not memory annulment occurred as a result of the procedure. If annulment occurred and only one mismatch is found in the procedure, then that one mismatch must have produced a PE experience (according to the PE inference principle defined earlier). If annulment did not occur, then that one mismatch did not produce a PE experience (also as per the PE inference principle). Comparing procedural mismatches that did and did not create PE experiences is crucially instructive. If annulment resulted and two or more mismatches occurred, the evaluation takes into account their comparative salience levels and their possibly differing specificity effects, in order to gauge their roles in producing the PE experience(s) that allowed annulment to result. That is the procedure applied in each of the study reviews in Appendices A and B.

For completeness, another type of PE, described in [78], should be recognized. When acquisition training ends when learning is not yet complete (not "asymptotic"), reactivation conditions can be fully identical to acquisition procedures (such as reinforced CS-US presentations) and yet still produce a PE experience and destabilization, because that reactivation functions as further training that increases certainty relative to the uncertainty of the incomplete target learning. This is described (p. 417) as "the requirement for memory updating to optimise further the predictive accuracy of the memory." In such cases, the reactivation experience contains a mismatch not in the procedure, i.e., not in what observably happens, but in the subjective experience of stronger certainty than exists in the target memory. Here again the distinction between external procedure and internal experience is important. This non-procedural type of mismatch and PE can be identified not from the comparison of acquisition and reactivation procedures, but rather from the responses measured for each successive acquisition trial, showing whether asymptotic learning was achieved. None of the studies reviewed in Appendices A or B involve this type of mismatch and PE.

3.2.3 PE as is defined to be a subjective experience that may or may not occur in response to a particular mismatch defined as a feature of the external procedure

The terms "memory mismatch" and "prediction error" have been used interchangeably as synonyms in MR literature, but here a fundamental distinction between them is introduced: A reactivation mismatch is defined to be an objective feature of the external experimental procedure, namely a difference (discrepancy) between the conditions of original memory acquisition and the reactivation and/or post-reactivation conditions, whereas a PE experience is defined to be a subjective event that may or may not occur internally in response to a particular procedural mismatch. This is yet another form of the distinction between procedures and experiences. In response to the same reactivation mismatch, some subjects may have a PE experience and some may not. Destabilization results only from a mismatch that creates a PE experience, as defined here, and does not result from a mismatch that does not create a PE experience. Only PE experiences cause memory destabilization, in this conceptualization.

3.2.4 Mismatches that are sub-salient and super-salient do not create a PE experience

A sub-salient mismatch does not cause a PE experience because the mismatch is too small or indistinct relative to what is expected according to the target memory. For example, the target memory contains contingency uncertainty (such as when partial reinforcement during acquisition makes the occurrence of the unconditioned stimulus (US) unpredictable) and the mismatch during reactivation is not large enough to differ unambiguously from that uncertainty. The subject may even notice the mismatch but feel unsure or doubtful about whether a discrepancy actually occurred.

A super-salient mismatch is defined in this framework as one that is in some way so large, quantitatively or qualitatively, that the experience is regarded by the subject as having a different latent cause than was in effect during the acquisition of the target memory, i.e., as being generated by a fundamentally different condition of the world and therefore as being qualitatively unrelated and irrelevant to the target memory [120, 121, 122, 123]. Because the knowledge and expectations in the target memory are then contextually irrelevant to the current, greatly mismatching experience, the knowledge and expectations in the target memory are not being violated by the current experience, so there is no PE experience generated and no destabilization of the encoding of the target memory. Consequently the current experience drives new, separate learning and the formation of a new, separate memory of the super-salient mismatch event, including its different latent cause.

As noted in the previous subsection, this is a new conceptualization of reactivation mismatches and PE experiences, differing from that of other authors who refer to all degrees of procedural mismatch as prediction error. For example, "large prediction error" has been used [124] to refer to mismatches too large to induce destabilization. However, that terminology is a conflation of concepts and a category error, because when the procedural mismatch is large enough to induce construal of a different latent cause rather than trigger destabilization, an experience of PE does not occur. That is really the core point of the "latent cause" concept. A prediction "error" exists only in relation to the originally construed latent cause, so as soon as perceived mismatch or discrepancy is large enough to induce construal of a different latent cause than that in the target memory, the target memory's predictions cease to be functionally relevant to the current mismatch event, and there is then no longer any "error," only new learning forming a new, separate memory.

In this framework, mismatches large enough to bring construal of a new latent cause do not create PE and are not labeled PE. Therefore, there is no such thing as a PE experience that is too strong to induce destabilization, because a mismatch ceases to create a PE experience when the mismatch is large enough to induce a new latent cause to be inferred. What can be too large to induce destabilization is a mismatch in the external reactivation or post-reactivation procedure, not an internal PE experience, because a too-large mismatch induces a change of latent cause, not a PE experience. As defined in this framework, every PE experience causes destabilization and does not change the inferred latent cause. If a PE experience has occurred, a destabilization has occurred. (The reverse principle defined above, that every destabilization is due to a PE experience, has the objective, empirical basis reviewed in Section 3.1 above.)

These considerations bear upon the findings of a sizable number of studies showing that reactivation mismatch has to be moderate, neither too mild nor too strong, in order to destabilize a memory [109, 110, 111, 112, 124, 113, 124, 20, 114, 115, 116]. An inverted-U graph of destabilization probability versus degree of reactivation mismatch has been described by [114]. Figure 1 updates that depiction according to the framework detailed above. Each of the indicated threshold levels of *degree of mismatch*, $\mathrm{DOM}_{\mathrm{PE}}$ and $\mathrm{DOM}_{\mathrm{LC}}$, depends on and varies with many parameters, including the strength of the acquisition training, the predictive uncertainty in the acquisition training (i.e., the reinforcement schedule), the specific feature(s) of the acquisition training being mismatched, and many subject-specific variables including level of trait anxiety, tolerance for uncertainty, internal style of contextual organization of experience, and allelic differences [34, 125].

In studies with asymptotic, 100% reinforcement during acquisition training, reactivation by a single unreinforced conditioned stimulus (CS) is a very strong mismatch because it is an absolute and total contradiction of the expected CS-US pairing, yet destabilization and memory annulment are observed to result [34, 110, 116, 84, 74]. This indicates that though the mismatch was very strong, it was perceived as having strong, clear relevance to the target memory and as having the same latent cause as in the acquisition training, therefore it created a PE experience. New latent causes are induced by mismatches that the subject discerns as indicating a different type of experience or phenomenon than was present in the acquisition training, and that qualitative discernment is not simply a matter of strength of mismatch.

This conceptualization of the creation of PE experiences and changes of latent cause differs sharply from that of Monfils and Holmes [105]. Despite the large body of empirical evidence showing that behavioral updating of emotional memory requires a PE experience that first destabilizes the memory (as reviewed above in section 3.1), these authors have interpreted an animal study by Gershman et al. [122] as indicating that "behavioural updating during reconsolidation does not require prediction error". That statement seems contradicted, however, by the PE experiences certainly produced in the post-reactivation procedure of [122] by presentations of an unreinforced CS that had 100% reinforcement during acquisition training one

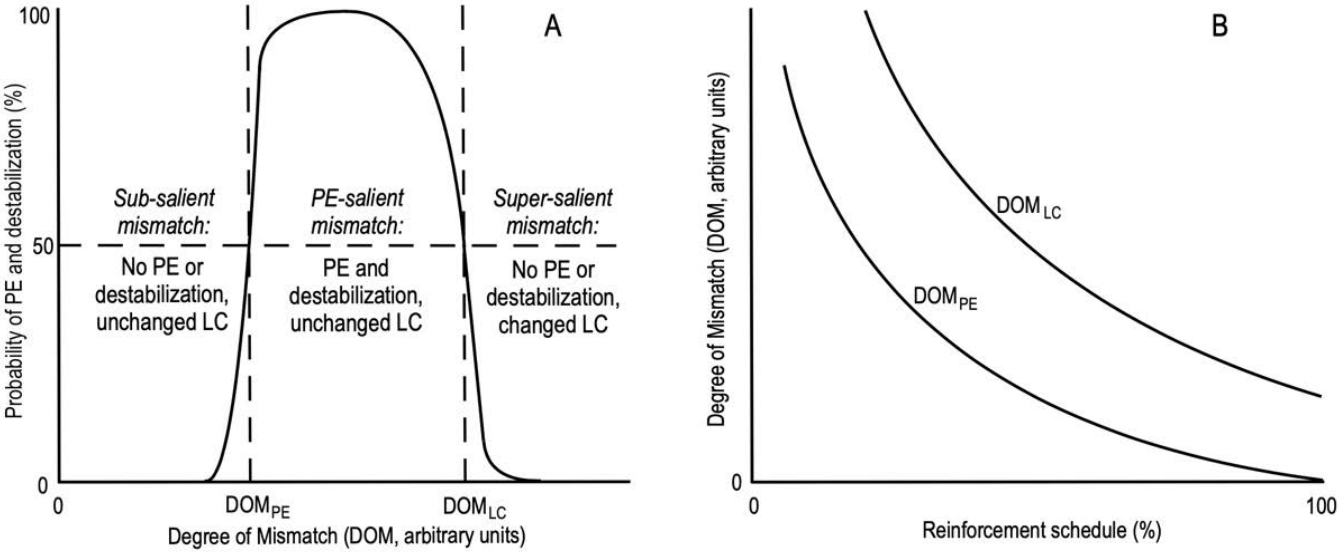

Figure 1: Schematic diagrams of (A) how the probability of a prediction error (PE) experience and memory destabilization putatively depends upon the degree of mismatch (DOM) between acquisition and reactivation conditions (LC = latent cause), and (B) how the two threshold levels of DOM vary with the target memory's degree of predictive uncertainty, indexed by the reinforcement schedule of the acquisition training: as reinforcement schedule increases to 100%, the target memory's predictive uncertainty decreases to zero. Asymptotic learning at acquisition is assumed.

day earlier. Nevertheless Monfils and Holmes assert that a PE experience is not necessary because "old memories are updated when new observations are inferred to be explained by the same latent causes" (p. 3). That seems to be an ad hoc assertion that if a new observation (i.e., a mismatch of what is expected according to existing memory) does not induce a new latent cause, then the new observation simply updates the old memory directly, without any PE experience occurring. However, unexpected "new observations" with unchanged latent cause are precisely what constitutes a PE experience. It is not apparent from their argument why such unexpected new observations would not create a PE experience. Merely asserting that a PE does not occur is not a sufficient or plausible analysis. The dichotomy set up by Monfils and Holmes, between PE experiences versus new observations with unchanged latent cause and no PE, seems to be a false dichotomy. In the framework delineated above, PE experiences are understood to be unexpected new post-reactivation perceptions with unchanged latent cause. In addition, whereas Monfils and Holmes maintain that the study by Gershman et al. supports their view that PE is not required for updating to occur, the conceptual framework provided by [122] allows that "one could modify the original fear memory if prediction errors were small or infrequent enough to not induce formation of a new memory, but still large enough to drive learning" (p. 2). That description, which clearly does not disavow PE experiences, is in agreement with the framework delineated above and corresponds to the middle section of Figure 1A.

3.3 PE evaluation framework applied to the retrieval-extinction protocol

The seminal study by Schiller et al. [6] first reported the lasting elimination of an acquired fear in humans as a result of MR-mediated behavioral updating. Due to the very strong clinical relevance of preventing the return of fear, the three-day protocol used in that study, known as retrieval-extinction or post-reactivation extinction, became a standard template for behavioral updating studies that many researchers subsequently have reused, with purposeful variations, for studying the behavioral unlearning and annulment of acquired emotional responses. This was the human application of a procedure that had previously been effective with rats [5] (and see [28] for the antecedents of this protocol in earlier research). In Tables 1 and 2, all but one [126] of the studies listed used variations of this protocol.

Schiller et al. [6] carried out two versions of the retrieval-extinction protocol, a between-subjects design (experiment 1) and a within-subjects design (experiment 2), both implemented on three consecutive days. Both designs begin with acquisition of a Pavlovian conditioned fear response on Day 1, with subjects viewing a computer screen and seeing presentations of either one (between-subjects) or two, separately presented (within-subjects) conditioned stimuli (CS+) that are paired on 38% of their presentations with an aversive unconditioned stimulus (US, electric shock) through attached equipment. Also separately presented on the screen is a different, neutral stimulus (CS–) that is never paired with the US. The CSs were differently colored squares. Fear responses were measured as the skin conductance response (SCR), which detects both negative and positive emotional responses. Other researchers have used the fear-potentiated startle (FPS) response, which measures amygdala-generated fear specifically.

On Day 2 of the protocol, one CS+ is presented once, unreinforced, as a reminder cue that reactivates the associated threat memory. Optionally, the neutral CS– may also be presented once. That is followed by a 10-min interval with an entertaining video displayed onscreen (a TV show episode was used in [6]), at the end of which the reactivated threat memory is in a destabilized condition if the protocol is having the intended effects. A standard extinction procedure is carried out next.

The post-reactivation procedure of a 10-min pause and extinction is found to result in no return of fear of the reminded CS+ when tested on Day 3 using standard techniques for re-evoking an extinguished fear response. Schiller et al. tested for spontaneous recovery via a re-extinction process (i.e., unreinforced presentations of CS+ and CS–). Other researchers have also tested for reinstatement by unsignalled presentations of the US followed by a re-extinction process, and/or have tested using reacquisition training.

In the between-subjects design with only one CS+, the group with no reminder before extinction on Day 2 shows return of fear of CS+ on Day 3. In the within-subjects design, subjects on Day 3 respond to the non-reminded CS+ with fear but show no fear of the reminded CS+. Schiller et al. also showed that fear is found to return on Day 3 even with CS+ reminder if the 10-min interval after CS+ reminder presentation is either eliminated (converting the entire Day 2 procedure to standard extinction) or is replaced by a 6-h interval (delaying the extinction training until the reminded CS+ threat memory is no longer destabilized).

It is clear from those results that the 10-min post-reactivation delay period is critically important for updating and annulment to result from the retrieval-extinction protocol. Therefore the 10-min interval is present in all subsequent human studies of behavioral updating of emotional memory (with significant variations of other parameters), i.e., all of the studies in Tables 1 and 2 except [126]. However, the attentional content of the 10-min interval has varied across studies to be either an onscreen video (TV episode, cartoon, nature scenes, or documentary of a sandcastle competition), magazines available for reading, or simply waiting/resting while remaining awake. All three categories of attentional content are present in both Tables 1 and 2, so the 10-min interval content cannot itself be the cause of replication or non-replication (although it could be a contributing factor influencing whether the active latent cause is maintained or changed, as discussed below in Section 3.3.1). The Table 1 (Table 2) percentages of studies are as follows: onscreen video 58% (64%), magazines 10.5% (22%), waiting 31.5% (14%). Each review in Appendices A and B indicates the particular content of the 10-min period.

3.3.1 PE and latent cause analysis of the retrieval-extinction protocol

In the large majority of retrieval-extinction studies, the Day 1 acquisition training was done with partial reinforcement (38% in the original version [6] and several others, and many others in the 40–80% range, as indicated in Tables 1 and 2). In consequence, on Day 2 the single unreinforced CS+ presentation is neither a procedural mismatch nor a PE experience. It is, rather, the singleness of that one CS+ presentation that mismatches the acquisition training and could create a PE experience, because the expectation in the CS+ threat memory is for a series of CS images to appear onscreen, separated by the same inter-trial interval as in the acquisition training. That PE experience would destabilize the CS+ threat memory, allowing it to be updated and nullified by the counter-learning experience in the subsequent extinction procedure.

However, whether that singleness of CS+ presentation actually does create a PE experience depends critically on the details of procedure immediately following offset (disappearance) of the CS+ reminder. Not seeing the expected series of CS+ images during the ensuing 10-min interval would register as a PE experience only if the latent cause is not changed by how the transition to the 10-min interval is presented to the subject. If the details of that transition do induce a change of latent cause, fundamentally changing the subjective context of what is being perceived, then no PE experience occurs, so no destabilization occurs and no updating results from the extinction training that follows the 10-min interval. Therefore, whether the study succeeds or fails to nullify the target memory is determined by the very brief section of procedure immediately following CS+ offset, i.e., the transition from CS+ offset to the 10-min interval. How that transition occurs determines whether the latent cause continues unchanged (resulting in a PE experience due to CS+ reminder singleness) or is immediately changed (precluding any PE experience). The following two sample scenarios illustrate those two possible outcomes.

The latent cause would persist unchanged and a PE experience would be created if CS+ offset is followed by a blank computer screen for the expected inter-trial interval and then a TV show episode or nature video appears for 10 min, with no communication to the subject from researchers about that transition and no disconnection of equipment from the subject's body. Under those conditions, subjects would regard the unexpected onscreen material as being part of the study, so the active latent cause would continue unchanged, maintaining the CS+ threat memory as the relevant expectational framework. The TV show or video appearing unexpectedly instead of another CS image would be experienced subjectively as the CS+ threat memory making a prediction error. A transition procedure of this type, leading subjects to regard the 10-min period as being part of the study, is described as follows in the study [127] reviewed in Appendix A.20: "A resting period of 10 min was inserted between reactivation and extinction. During the break, participants watched an excerpt from [the BBC documentary] Planet Earth. During all sessions of the experiment, the shock stimulator was set to the "on" position and the SCR was recorded continuously" (p. 6). Subjects remaining connected to the electronic equipment set to "on" during the 10-min delay presumably led them to regard the 10-min period as being part of the study, allowing the single CS+ reminder presentation to create a PE experience, destabilizing its associated threat memory. This study reported successful updating and nullification of threat memory.

In contrast, a transition after CS+ offset that subjects interpret to mean that the 10-min period is not a part of the study, inducing a change of latent cause and preventing a PE experience from occurring, is exemplified by this description from a study that reported a replication failure [50] (reviewed in Section B.11):

"After a single presentation of the [reminder CS], the computer screen went black, and the following instructions were given: "We now will have a ten-minute break. Here are some magazines that you may choose to read during the break. You do not have to read anything if you do not want to. The only thing that is important is that you take your break in the specified location where I take you. However, first, I will turn off and disconnect you from these devices". The subject was disconnected from the electrodes, the stimulator was set to the "Off" position, and the participant was taken to a waiting room with chairs and magazines. Once the ten-minute period was over, participants were returned to the testing room, and the following instructions were given: "We have finished the ten-minute break. I will now reconnect these devices, and we will complete the remainder of the session". Subjects were reconnected to the electrodes, and the stimulator was set to the "On" position. The session was resumed with the remaining presentations of the CSs (i.e., 14 CSa+, 15CSb+, and 14 CS-)." (pp. 129–130)

With the transition done in that manner, the subject regards the coming "break" as not being part of the study, creating a different active latent cause of the "break" than was just previously in effect during the study's CS+ reminder presentation. Therefore, immediately after CS+ offset, the CS+ threat memory is no longer the active expectational framework and whatever now happens, even if unexpected, cannot create a PE experience because now there is no memory-based expectation in effect. The 10-min period and its contents are now subjectively irrelevant to the CS+ memory and its latent cause, so the video presented or magazines offered do not have the meaning of PE in relation to the target CS+ threat memory, which therefore is not destabilized, precluding later updating. Notably, this explanation of non-replication applies even to the study [52], reviewed in Section B.12, that importantly attempted an exact, registered, high-powered replication of the original study by Schiller et al. [6].

Thus, according to this framework of PE evaluation, the procedural details of the transition from CS+ offset to the 10-min interval control whether any PE experience is created in the retrieval-extinction protocol with partial reinforcement. The details tend to be unique in each study and must be closely evaluated in each case. For example, if CS+ offset were to be followed by a blank computer screen for 1 min, which is a multiple of about 5 times longer than the expected inter-trial interval learned during acquisition, a (temporal) PE experience would presumably be created by the end of that 1-min period, destabilizing the CS+ threat memory even if next the subject were told, “Next is a 10-minute break, and a TV episode will play for you to enjoy, and no shocks will occur.” That is a testable prediction of this framework of PE evaluation. In that case, a PE experience would be created not by the 10-min interval, but by the single CS+ presentation not being followed by other CSs for 1 min.

The entire outcome of a retrieval-extinction study with partial reinforcement hinges on this phenomenology at this point of transition from the CS+ presentation to the 10-min interval in the Day 2 procedure, in this analysis. (In the case of 100% reinforcement in acquisition, as used by three studies in Table 1, the single unreinforced CS+ at Day 2 reactivation does reliably create a PE experience.) As a rule, this critical transition has been overlooked both by study authors who have not described these procedural details and by authors of numerous review articles discussing the possible causes of non-replications of behavioral updating. Most authors, like Schiller et al. [6], indicate only the content of the 10-min period and very little or nothing about the transition. Exceptions are the explicit transition details provided by three studies in Table 1 [117, 127, 128] (reviewed in sections A.16, A.17, and A.20) and by two studies in Table 2 [50, 129] (reviewed in sections B.11 and B.13). In each of those five cases, the transition details can explain why the study produced a replication success or failure.

The possible governing role of the transition to the 10-min interval has been identified previously only twice [28, 130], to the author’s best knowledge. Zuccolo and Hunziker [28] stated, “Several of these reports describe disconnecting participants from US-delivering devices after the presentation of the retrieval cue and reconnecting the devices just before extinction training. Therefore, it is logical to suppose that events occurring during this interval might be relevant for changing the effects of the intervention.” They further refer to that as “changes in experimental context”, which is conceptualized in the present article as a change of latent cause, and they conclude this topic by stating (p. 51), “it is necessary to consider the role of verbal instruction in prediction error.”

In each study in Appendices A and B, the transition is analyzed for the presence or absence of a PE experience, to the extent possible using the procedural information provided by study authors, which varies widely. This transition is in that way found to be a serious candidate cause of non-replication in 11 of the 14 non-replication studies, as indicated in Table 2.

In addition to the manner of transition from Day 2 reactivation to the 10-min delay, there are three other factors that can influence whether the latent cause construed in Day 1 acquisition procedures remains in effect through Day 2 reactivation and post-reactivation procedures sufficiently for a PE experience to be created. All four factors are listed here:

- how the transition from CS+ offset to the 10-min interval is communicated to subjects
- the content of the 10-min interval in relation to acquisition memory content
- the degree of uncertainty of target memory latent cause
- individual differences among subjects

These four factors are important determinants of the outcome of the reactivation-extinction protocol. The transition factor has been described above. Each of the other three factors is explained below:

- **The content of the 10-min interval in relation to acquisition memory content**: In the original study [6] and in several subsequent studies, during acquisition on Day 1 colored squares on a computer screen served as the CSs. If the 10-min video on Day 2 were to begin immediately after reminder CS+ offset and show similar geometric forms drifting around, the video would be perceived as relevant to the acquisition training and would not induce a content-driven change of latent cause for most if not all subjects. Rather, a PE experience would be produced by the novel, unexpected sight of the video following the CS presentation. In contrast, a video of angry men fighting would be more likely to induce a change of latent cause for most subjects. However, if the CSs during Day 1 acquisition had been images of men’s aggressive or angry faces (as in experiment 1 of [131]), a Day 2 video of angry men fighting would seem relevant to the acquisition training and would therefore tend to produce a PE experience. Thus, the same video could create PE in one study but induce a change of latent cause in another study. The degree of mismatch between acquisition training content and 10-min interval content can in that way influence whether or not destabilization and updating result. However, this effect is presumably moderated by the effect, described above, of how the transition to the 10-min period is communicated to subjects. For example, if the researchers inform subjects that the 10-min period as an important component of the study, it probably would not induce a change of latent cause even if video content seems utterly unrelated to anything in the study.

- **The degree of uncertainty of target memory latent cause:** 100% reinforcement of the CS-US association during Day 1 acquisition training creates maximum certainty of latent cause. Partial reinforcement produces uncertainty of latent cause. Smaller percentage of reinforcement produces larger uncertainty of latent cause. After acquisition with partial reinforcement, Day 2 reactivation by an unreinforced CS presentation is then somewhat ambiguous as to whether its latent cause is the same or different from the latent cause of the acquisition training, further increasing the uncertainty of the currently active latent cause. Then the video appears. If at that point the target memory's latent cause has strong uncertainty, then whether the video does or does not have the same latent cause is also uncertain even if video content has no obvious relevance to acquisition training content, and the video is perceived as not definitely irrelevant to the current latent cause, i.e., possibly relevant to it, which preserves the latent cause. Consequently the mismatch of seeing the unexpected video generates a PE experience and destabilizes the target learning. This effect could be a significant contributor to the success of retrieval-extinction studies with a low level (38%) of reinforcement during acquisition (see Table 1). If, however, latent cause certainty is strong when the video appears, then the video's irrelevance to that latent cause is clearly apparent (unless video content is strongly similar or relevant to acquisition training content), which would tend to induce a change of latent cause, preventing destabilization.
- **Individual differences among subjects:** Differences in the threshold mismatch for construing a change of latent cause could be caused by differences in various psychological characteristics, such as trait anxiety, tolerance for uncertainty, and internal style of contextual organization of experience. In addition, differences in genetic alleles have been shown to correlate with significant differences in memory processes [34, 125].

The above considerations are applied where relevant in the reviews of individual studies in Appendices A and B.

The uniqueness and specificity of the PE evaluation of each study are illustrated well by a comparison of results reported by Sevenster et al. [116] and Chen et al. [110], both of which were human studies (see Section A.18 for a review of the latter study). Sevenster et al. found that threat memory destabilization occurred, allowing annulment via propranolol, from a reactivation consisting of one presentation of a mismatch of the expected CS-US contingency, but did not occur from a reactivation consisting of a series of three presentations of that same mismatch. In contrast, Chen et al. found that for a threat memory having strong uncertainty in US timing relative to the CS, destabilization did not occur when reactivation consisted of one presentation of a mismatch of the expected CS-US contingency, but did occur from reactivation consisting of two presentations of the same mismatch, allowing annulment via behavioral counter-learning. Both of those seemingly opposite results are fully consistent with the PE evaluation principles defined above and illustrate that it is the unique relationship between the acquisition conditions and the reactivation and post-reactivation conditions that determines whether a particular reactivation or post-reactivation mismatch creates a PE experience, destabilizing the target memory. For the target memory of Sevenster et al., reactivation with one mismatch sufficed to create PE (middle section of Figure 1A) but a series of 3 mismatches was a large enough net mismatch to induce instead a change of latent cause (right-hand section of Figure 1A). For the high-uncertainty target memory of Chen et al., reactivation with one mismatch was not sufficiently salient to create a PE experience (left-hand section of Figure 1A), but a series of 2 mismatches was sufficiently salient (middle section of Figure 1A).

3.3.2 PE specificity destabilizes specific components of the acquisition memory, explaining idiosyncrasies in return-of-fear test results

In many of the studies listed in Tables 1 and 2, the return-of-fear test data contains a distinct, anomalous feature, for which the authors speculate about various possible causes. These unexpected, difficult-to-explain features include:

- fear in response to CS– on Day 3, such as reported in [117] and reviewed in Section A.16;
- CS+ fear response is eliminated at the end of the Day 2 procedure but returns 3 h later, then is again absent at 24 h, as reported in [127] and reviewed in Section A.20;
- failure to induce CS+ threat memory updating but success inducing episodic memory updating by the same reminder in the same procedure, as reported in [132] and reviewed in Section B.8;
- gender difference in response to reinstatement, with males showing return of fear and females showing no return of fear, as reported in [133] and reviewed in Section A.19.

These diverse phenomena are the strongest tests of the capability of the proposed trial framework of PE evaluation to provide cogent conceptual explanations for all findings in studies of behavioral updating of human emotional memory. As delineated in each study's review in Appendix A or B, each of these phenomena is explainable in detail by the principle of PE specificity, which is a critically important component of this unified and unifying framework.

The specificity of memory destabilization and updating is essential for clinical application and has been demonstrated in several studies [56, 128, 134, 36, 38, 6, 33]. In the experimental paradigms demonstrating specificity, a conditioned response is created for two different CSs or two different USs, and then memory reactivation by one CS or one US is shown to destabilize only the memory and memory linkages of that specific stimulus, leaving memory of the other CS or US unaffected. That has been conceptualized by researchers as cue specificity, but its relationship to the specificity of PE phenomenology has not been considered.

As noted earlier, in any given study, the acquisition training creates a memory with numerous components, typically including the percepts of each CS (whether visual image, sound, or other), the duration of each CS presentation and of the inter-trial interval, the sensation and intensity level of the US, the serial, repetitive pattern of CS presentations, the total number of CS presentations, the degree of uncertainty of CS-US pairing, the latent cause, and the context. What is being introduced here is the concept of PE specificity, consisting of the recognition that (a) a PE experience is, as a rule, the subject's perception of a discrepancy in how one or more of those specific components of the acquisition memory compares with current experience during the reactivation and post-reactivation procedure, and (b) only that specific, discrepant component of the acquisition memory is destabilized by the PE experience and is subsequently updated by an extinction procedure, leaving other components of the memory intact, unchanged, and functional.

Therefore, different reminders that create different PE experiences can produce very different results of memory updating, and understanding experimental results requires a detailed analysis of the PE specificity of the procedure. For example, a discrepancy in the CS+ reminder image (relative to the CS+ image presented during acquisition training) can create a PE experience that destabilizes only the CS+ visual image component of the acquisition memory, leaving other components stable, such as the CS+/US contingency memory.

For studies that used presentation of the US as the Day 2 reminder, application of the PE specificity principle reveals rich structure. Only two human studies of the retrieval-extinction protocol have used US reactivation, but both ([134] and [74], reviewed in Sections A.5 and A.11, respectively) produced updating and annulment of threat memory, suggesting the possibility that US reactivation may be more reliable for causing threat memory destabilization, which implies greater reliability for creating a PE experience. In animal studies, no failures to achieve destabilization with US reactivation have been reported, to the author's best knowledge, and several animal studies with US reactivation have shown that specific sensory associations are encoded separately and can be selectively destabilized [79, 95, 135], leading some authors [135] to comment (p. 537), "The selectivity of reconsolidation processes seems to protect the global integrity of memories."

Why US reactivation could have heightened effectiveness becomes apparent from applying the PE specificity principle to the two human studies with US reactivation (74, 134). During acquisition on Day 1, the US, an electric shock, was always closely preceded by a CS+ and occurred exactly at CS+ offset. Therefore, presentation of the US alone on Day 2 mismatches the target memory's knowledge of when the US occurs, presumably creating an immediate PE experience that destabilizes the target CS+/US contingency and threat memories. In addition, the intensity of the US shock was reduced by 50% for reactivation in both studies. The significantly reduced intensity is also a specific, salient mismatch and violation of expectation, which may reasonably be assumed to create a separate, additional PE experience bearing upon the US sensory memory. The PE specificity principle calls attention to the possibility that different components of the multi-component acquisition memory can be selectively destabilized by different PE experiences. That is, the electric shock sensation is a different component of memory than the representation of the CS+/US contingency relationship, and the PE specificity principle posits that those different components can be differentially destabilized by component-specific mismatches. Qualitatively different components of the acquisition memory may be encoded in different, but linked, memory systems, allowing their separate, differential destabilization.

Further, the CS+/US contingency relationship has its own components, including not only the obvious elements of sequence and timing, but also the less obvious element of causality: the US shock always occurring at CS+ offset during acquisition appears to mean that the CS+ is the cause of the US and makes US occurrence predictable. That model of causality is itself a component of the target threat memory, and it is mismatched saliently by the US occurring alone for reactivation on Day 2, disconfirming the CS+ as the cause of the US and removing the predictability of the US. It is then suddenly apparent that the US has its own power to occur, independently of the CS+, which may somewhat reduce the causal role and threat potency of the CS+. Such large qualitative changes could conceivably induce a change of latent cause instead of producing a PE experience, but the fact that destabilization and updating do result can only mean that these changes register as a PE experience, not as a change of latent cause.

Thus, US reactivation may have higher reliability than CS+ reactivation for producing destabilization and updating because it very probably creates multiple specific PE experiences by itself, even before an additional PE experience may be created by the post-reactivation procedure, such as the unexpected appearance onscreen of a nature video that plays for 10 min. This PE evaluation of US reactivation makes the testable prediction that the 10-min delay period is unnecessary for destabilization to occur, in sharp contrast with CS+ reactivation after partial reinforcement acquisition. Eliminating the 10-min period and beginning the extinction procedure 1 or 2 min after US reactivation (enough time for launching the neurochemical mechanisms of destabilization) should result in successful updating. That would establish that a 10-min delay is not an inherent requirement for updating and nullifying a fear memory, as it appeared to be in studies such as that of Schiller et al. [6], where the 10-min delay happened to be the only source of a PE experience. When another source of PE is also present, the 10-min delay will prove to be unnecessary, according to this prediction of the PE framework proposed here.

The proposed model of PE specificity is also amenable to empirical testing by further elaborating the use of compound CSs (such as in the study [113] reviewed in Section A.12) or compound USs (such as in the study [136] reviewed in Section A.8), and testing various reminder cues that contain a range of different subsets of the component elements of the compound CS or US. For some designs of that type, a PE specificity analysis could be made in advance and pre-registered, predicting the differential results of different return-of-fear tests (spontaneous recovery, reinstatement, reacquisition). Other designs of that type could probe finely across a continuum of reminder variations in order to reveal nuances of mismatch, PE, latent cause, and destabilization phenomenology, refining our knowledge of PE phenomenology and the PE specificity model. For example, CSs consisting of three visual components each were used in [113], and a reminder with two of the three elements was the smallest degree of mismatch relative to acquisition conditions in that study, but a continuum of smaller degrees of visual image mismatch could be tested by retaining all three visual elements in the reminder and varying the size, spatial orientation, shape or color of one or more of them. If the three-element reminder is presented with the added small image of an ant near one of the geometric shapes, would a PE experience and destabilization result?

Yet other possibilities are suggested by the review in Section B.8 of the study in [132], where a PE specificity analysis finds that the Day 2 reactivation image failed to create any PE experience in relation to the generalized threat memory that was acquired on Day 1, resulting in failure to prevent return of fear. However, that same reactivation image did create a PE experience in relation to episodic visual memory, resulting in MR-induced episodic memory enhancements relative to the no-reminder group. That example points to a universe of possibilities for testing PE specificity by designing studies in which different subject groups have different reminders designed to differentially destabilize different components of the acquisition memory.

4. Significant patterns in successful human behavioral annulment studies

4.1 Summary of research findings identified in Appendix A

Appendix A provides a review of each of the 20 successful human behavioral updating and annulment studies listed in Table 1. This collection of reviews is the most comprehensive compendium to date of human studies that have reported full annulment of an emotional learning by MR behavioral updating. Each study is examined through both the lens of concrete procedures and the lens of subjective experiences. Each review has the minimum level of technical detail necessary for comparing the studies' primary procedural features that directly implemented their experimental strategies and designs. Not included in the accounts are elements involved in empirical pragmatics and accuracy, such as the number of subjects, information on habituation trials, or statistics. Criteria for excluding studies from Table 1 are defined in Section 1.

From the reviews in Appendix A, several significant patterns emerge unambiguously and uniformly as being the core findings of research on the unlearning and annulment of human emotional learnings:

- Concrete procedures varied widely, yet each study's unique procedure caused subjects to have the same, invariant set of three experiences: target memory in reactivated state, an experience of target memory prediction error (PE) during that reactivation (for destabilizing the target memory), and a counter-learning experience contradicting the target memory's knowledge and expectations (for driving unlearning, updating, replacement, and annulment of the target memory).
- The largest variation of procedure was the study [126] reviewed in Section A.7, which did not use the retrieval-extinction protocol but nevertheless created the same three experiences, thereby demonstrating that what matters for annulment of emotional memory is that set of experiences, not any particular concrete procedure for creating those experiences. It is therefore that set of three experiences, not any particular procedure, that warrants being identified as the mediator that recruits the MR mechanism for the unlearning and annulment of emotional memory.
- Also warranted is a name for that critical set of three experiences: hereafter, the unlearning trifecta. The current review establishes strong empirical support for the hypothesis that the human brain unlearns and nullifies an emotional learning via MR in response

to the unlearning trifecta. That support consists of 20 confirming studies (Table 1) combined with analyses indicating that the 14 non-replications (Table 2) failed to fulfill the unlearning trifecta, as explained in Section 5 below. As discussed in [13] in the context of mechanism of change in psychotherapy, the unlearning trifecta is the mediator, or operational condition, that engages the internal mechanism of MR to produce the outcome of transformational change, i.e., full annulment of a specific emotional learning and its behavioral, somatic, and state-of-mind manifestations.

- The strength of the unlearning and annulment effect has been emphasized by some study authors. For example, the study [136] reviewed in Section A.8 compared the responses of adults and adolescents, while noting that extinction procedures had previously been found to be less effective with adolescents than for both younger and older age groups. These researchers conclude (p. 3), "Participants who were reminded of the conditioned stimulus 10 minutes prior to extinction showed no recovery of fear 24 hours later. Conditioned fear did not return in the reconsolidation update group after reinstatement, even in our adolescents who showed diminished within-session extinction [relative to adults], highlighting the robustness of the effect." For both adolescent and adult reconsolidation groups, the level of fear after reinstatement on Day 3 was actually lower than the level of fear at the end of extinction on Day 2, which is a decisive nullification of the acquired fear response.
- Annulment of the target emotional memory is long-lasting, with annulment shown to persist out to 18 months [6, 35, 61, 134] (reviewed in Sections A.1, A.3, and A.5). A review by Phelps and Hofmann [137] notes that the experimentally documented disappearance of acquired fear responses in humans persisted for "at least a year—consistent with a permanent alteration of the threat memory." This means that "the changes resulting from MR behavioral updating are a 'new normal' that requires no effort or maintenance procedures of any kind for them to persist; this is the basis for describing the changes resulting from MR as 'enduring' or 'permanent' and 'effortless' to maintain." [10]
- Annulment was achieved for emotional learnings other than threat memory, significantly more complex than classical conditioning, and more relevant to clinical memory targets [126] (reviewed in Section A.7).
- Principled analysis and specific identification of PE experiences in a study's procedure is both feasible and necessary for understanding experimental results (discussed above in Section 3.1 and demonstrated for each study uniquely throughout the reviews in Appendices A and B).
- A procedure other than extinction training was used to create the counter-learning experience, namely change of contingency [126]. Likewise, in animal studies annulment of emotional memory has resulted from counter-learning other than extinction, namely hedonic valence reversal learning [138, 139, 140] and a gradually less frequent reinforcement training [122]. The conclusion from one animal study [22] was that "the updating of appetitive CS-US associations underpinning conditioned responding in manners other than extinction training is likely achieved through memory reconsolidation." Thus the experience of counter-learning during the reconsolidation window is not restricted to any particular procedure.
- The successful use of extinction training (a series of non-occurrences of an expected event) as the counter-learning experience in the retrieval-extinction protocol shows, importantly, that a destabilized emotional memory can be disconfirmed, unlearned, and annulled through MR by an experience that is in itself non-emotional by everyday standards, such as a yellow square appearing on a computer screen and then disappearing after several seconds with nothing else happening. A clinical case example in the next section demonstrates this phenomenon even more strongly.

Thus, the review here of 20 successful laboratory studies of human emotional memory annulment leads to understanding the behavioral updating process as consisting of a set of critical subjective experiences that drive it. It is worth noting that even if this review excluded the original study by Schiller et al. [6], as has been recommended based on a critique of that study's data processing procedures [37], the above summary of research findings and the seminal value and influence of that original study would be unchanged.

4.2 An explanation of the extinction-retrieval protocol in terms of MR

Understanding the operation of the retrieval-extinction protocol in terms of subjective experiences generates a new explanation for perhaps the most controversial observation in the field of MR behavioral updating research, namely the finding first in animal studies [75, 76, 77] and then in a human study [127] (reviewed in Section A.20) that the retrieval-extinction procedure and the reversed procedure of extinction-retrieval equally diminished the target emotional memory after 24 h. That finding has widely been regarded as possibly indicating that the retrieval-extinction protocol does not engage the MR mechanism after all, because "Reconsolidation theory would posit that retrieval must come before extinction for the procedure to impair reinstatement" [69] and "it is difficult to reconcile how reconsolidation can be initiated when the reminder trial follows, rather than precedes, extinction learning" [73]. Similarly, "When reactivation follows extinction, reconsolidation processes cannot truly be involved, so these findings question the logic and mechanism of reactivation-extinction in general" [52] (and see also [141]).

However, those are procedure-oriented analyses of the effects and operation of extinction-retrieval. In contrast, an experience-oriented analysis has been proposed [47] that is consistent with MR. A further-developed version of that analysis is as follows. In this view, the critical condition for updating and unlearning to occur is the concurrent experiencing of what is known according to the reactivated, destabilized target learning and what is known according to the freshly active counter-learning experience of extinction. It has been shown that the engram encoding a new learning can actively interact with a subsequent new learning that is created up to 5 h later, but not 6 h later [142, 143]. Thus, if target memory reactivation plus destabilization occurs and an extinction learning occurs and both are active concurrently, regardless of the order in which they are activated, the two mutually incompatible, coactivated knowledges generate PE and drive updating, with the older memory's pre-existing knowledge being revised by the new memory's new knowledge. Therefore, the fact that unlearning and annulment result from both retrieval-extinction and extinction-retrieval is fully compatible with the MR mechanism.

Beyond that general analysis, the details of the MR phenomenology of the extinction-retrieval protocol are somewhat more complex. A proposed account of that phenomenology is presented here as an exercise within this article's overall exploration of the question: Is it possible to account for all experimental observations of MR in terms of the requirement of a PE experience for initiating the MR process? The following extinction-retrieval scenario would answer that question in the affirmative.

How the target fear memory becomes destabilized and updated by the extinction-retrieval procedure is the first focus of this account. The appearance of the first CS+ at the start of the extinction training on Day 2 reactivates the acquisition threat memory, in which the CS+ is known to (sometimes) bring pain and therefore evokes a fear response. The extinction training then proceeds and produces a new memory engram, separate and distinct from the acquisition memory, as is empirically well established [144, 145]. During extinction the CSs appear regularly every several seconds, numerous times, but finally another CS does not appear after the expected number of seconds, rather a 10-min interval passes and then an unreinforced CS+ appears once. If the CS+ threat memory was acquired with partial reinforcement, a PE experience is not created by that single CS+ being unreinforced. However, never before has a 10-min inter-trial interval been experienced, so a temporal PE experience occurs, destabilizing the engram of the acquisition threat memory of CS+. Then, no CS follows. The acquisition threat memory is now both activated and destabilized in the presence of the new and still fully activated extinction memory consisting of the contradictory knowledge of the harmlessness of CS+. Therefore, updating and unlearning of the acquisition threat memory of CS+ now occurs immediately.

The next observations to be explained in terms of MR are these, again in the human study [127]: At 3 h or 12 h after the Day 2 behavioral unlearning procedure, fear memory testing (by two unreinforced CS+ presentations) showed return of fear for retrieval-extinction but not for extinction-retrieval, but no return of fear after 24 h and sleep for both procedures (tested for both spontaneous recovery and reinstatement). An animal study [77] had produced quite similar results. There was also no return of fear at 12 h with retrieval-extinction if sleep occurred during those 12 h.

The full annulment of fear responding at 24 h strongly suggests that both procedures recruited the MR mechanism. In order to consider how MR could be the operative mechanism in each of those procedures and yet produce their different effects, it is necessary to take account of specifically how retrieval-extinction and extinction-retrieval involve different processes of formation of contradictory knowledge through counter-learning: In retrieval-extinction, first the CS+ threat memory is destabilized and then the extinction procedure only gradually builds up a contradictory knowledge over time. In extinction-retrieval, first extinction creates a fully formed, full-strength, activated contradictory knowledge, and only then does the CS+ threat memory become destabilized (by the putative process described above). Those quite different sequences suggest the possibility that retrieval-extinction and extinction-retrieval might induce different types of updating, resulting in their different effects later on Day 2. Conjecturing the following two differing phenomenologies of updating seems necessary to explain the experimental observations:

- In retrieval-extinction, the learning being created by the first two trials of the extinction training is at first consistent with the acquisition threat memory of partial reinforcement, and then only gradually does the contradictory, disconfirming quality of the extinction training become apparent and finally stark. Because the destabilized threat memory is exposed initially to new learning that is consistent with it, updating occurs at first through a conjoining or integration of the extinction training's slowly developing engram with the threat memory engram. The latter remains intact though conjoined, so it continues to respond, generating fear, later at 3 h and 12 h. Human MR studies have reported such conjoining of old and new memory contents [146, 147]. However, during the next sleep, the updating process continues, transforming the initially conjoined engrams into a unified synthesis by revising the encoding of the older threat memory engram according to the newer no-threat engram. Then the threat memory no longer responds at 24 h (or at 12 h if sleep has already occurred), as observed.

- In extinction-retrieval, first the counter-learning is completed, so the contradictory knowledge is already fully developed and activated when the target threat memory then becomes destabilized. Destabilization in the presence of decisive disconfirmation produces immediate, complete updating and unlearning of the threat memory and a replacement of its engram by that of the no-threat memory, so no fear response is seen at 3 h or 12 h.

Thus it appears logically feasible that the extinction-retrieval protocol could be explainable in terms of MR. To summarize the bimodal updating phenomenology being conjectured here: If an emotional memory interacts immediately upon destabilization with a strong, decisive contradictory knowledge, it is directly and immediately rewritten and nullified, but if it is coactivated with a gradually forming contradictory knowledge that is at first not yet saliently contradictory, updating occurs in two phases, beginning with a phase of conjoining or integration of the two distinct engrams. That allows the target memory to maintain responsiveness until the first sleep, during which a second phase of full updating transforms the target memory's engram, nullifying it.

Regardless of what the true mnemonic processes of the retrieval-extinction and extinction-retrieval protocols prove to be, the fact that both result in no spontaneous recovery or reinstatement of fear after 24 h [127] puts firm ground beneath the hypothesis that each protocol recruits the MR mechanism (but see [127] for a different interpretation and discussion of other possibilities). A behavioral test of the above conjecture would be to revise the retrieval-extinction protocol of [127] to have 100% reinforcement in the Day 1 acquisition training, so that the extinction procedure on Day 2 rapidly creates decisive contradictory knowledge from the first trial. If updating is bimodal as proposed above, making the counter-learning abrupt instead of gradual would eliminate the return of fear at 3 h and 12 h.

Regarding the experience-oriented account of extinction-retrieval, Ecker [47] stated, "The idea that counter-learning could precede destabilization and still be effective for erasure ceases to seem counter-intuitive with recognition that for updating, the brain requires only a juxtaposition (concurrence, simultaneity) of the two experiences, regardless of which of the two is activated first. Each of those two experiences has an extended though limited duration of activation, which is why their juxtaposition is possible. …The observations of Baker et al. (2013) [75] and Millan et al. (2013) [76] can therefore be seen as supporting the experience-oriented interpretation of reconsolidation phenomena rather than as indicating non-recruitment of reconsolidation by the post-retrieval extinction protocol."

5. Significant patterns in unsuccessful human behavioral annulment studies

In all of the non-replication studies in Table 2, the target threat memory was created using 50 to 100% reinforcement, considerably higher than the 38% reinforcement of Schiller et al. [6]. That could seem to imply that stronger threat memory may be responsible for the non-replications to a significant degree. However, 12 of the 20 successful behavioral annulment studies in Table 1 also had 50% to 100% reinforcement, so that cannot be the main explanation. The same point is made in a review [28] that shows that "there is no apparent relation between reinforcement rate or number of CS-US pairings and fear reduction after PRE [post-reactivation extinction]" (p. 48). Those authors also amplify the argument made by other researchers [131] refuting the assumption that higher reinforcement rates produce stronger conditioning, citing empirical evidence that stronger conditioning results rather from "increasing the number of CS-US pairings and not by manipulating the reinforcement rate" (p. 48).

The present review suggests the possibility that each of the non-replications to date (i.e., all studies in Table 2) can be accounted for, cogently and testably, as being caused by an absence of PE experience as defined in Section 3 and Figure 1A, resulting in no destabilization of the target memory and therefore no updating or annulment.

In Appendix B, the PE evaluation of each study in Table 2 arrives at that conclusion. The studies were reported with varying degrees of detail for the procedural elements that are involved in evaluating the presence or absence of PE experiences, resulting in evaluations that range from confident to speculative. However, certainty is not needed for exploring the possibility that all studies might be explainable by a consistent, principled evaluation of PE.

Across the reviews and evaluations in Appendix B, the absence of a PE experience has several different identified causes, which are listed in Table 2 and are summarized here:

- With partial reinforcement at Day 1 acquisition in 13 of the 14 studies, the Day 2 CS reactivation did not create a PE experience by being unreinforced. (Note that none of the non-replication studies used the US for reactivation, which appears to be a reliable creator of PE, as discussed in Section 3.3.2 above and as suggested by the two US reinforcement studies in Table 1 [74, 134]. In these 13 partial reinforcement studies, whether or not a PE experience would occur on Day 2 was therefore determined entirely by the post-reactivation procedure immediately following reminder offset (disappearance), namely the transition to the 10-min interval between offset and the beginning of extinction training.
- In 11 of the 13 partial reinforcement studies, it is probable or definite that subjects were led to view and understand the 10-min period as being a "break," i.e., not part of the study. As discussed in Section 3.3.1, that would induce a new latent cause, preventing the PE experience(s) that would otherwise have been created.

- In the one study in Table 2 with 100% reinforcement [148] (reviewed in Section B.5), it is probable that reactivation mismatch was too weak to create a PE experience, based on comparison with another 100% reinforcement study [110] (reviewed in Section A.18) that used various reactivation conditions having different degrees of mismatch, only the strongest of which produced updating and annulment.
- In one study, the long duration of reactivation was very probably a too-large mismatch that induced a new latent cause rather than a PE experience [149] (reviewed in Section B.4), a conclusion supported by direct comparison to one of the successful studies [150] (reviewed in Section A.13).

All of the above PE analyses and conclusions are empirically testable. For each study that failed to replicate the updating and annulment effect, a specific change of procedure that should produce replication is identified in the study's review in Appendix B.

The numerous replication failures have widely been interpreted as meaning that behavioral updating has a limited range of applicability due to having complex, delicate boundary conditions that make it a fragile phenomenon [28, 55, 131, 149]. For example, Chalkia et al. [52] came to the conclusion, based on their replication failure, that (p. 496) "The results of the present study, along with the mixed findings in the literature…, give cause to question whether there is robust evidence that reactivation-extinction prevents the return of fear in humans."

However, if the PE analyses in Appendix B and summarized above were to be tested empirically and supported, the implication would be that no limits of behavioral updating have yet been found because all non-replications were due to specific, avoidable experimental conditions that caused mismatches to be too large or too small to create a PE experience. The updating and annulment effect would then be recognized as being reliable and sturdy rather than elusive and fragile, requiring only a mismatch of the threat or contingency memory that creates a PE experience, as represented in the middle section of Figure 1A, and as the 20 studies in Table 1 apparently have done.

Understood as being caused by the absence of a PE experience, the non-replication studies could serve to help the MR research field to arrive at a more unified recognition of the PE requirement and an evolving, more nuanced model of PE phenomenology. This review's trial framework of PE evaluation, mapped out in Section 3 and Figure 1, is presented in that spirit.

6. Clinical translation

MR researchers Elsey and Kindt [56] acknowledge that "there are significant limitations to experimental research, and ultimately only attempts at treatment can reveal the utility of a reconsolidation-based approach." This section explores the clinical translation of the empirical findings listed above in Section 4, specifically the set of three experiences that resulted, in every MR study in Table 1, in unlearning and annulment of emotional memory through the mechanism of MR, and which collectively are here termed the unlearning trifecta:

- target memory in reactivated state
- an experience of target memory prediction error (PE) during that reactivation (for destabilizing the target memory)
- a counter-learning experience contradicting the target memory's knowledge and expectations (for driving unlearning, updating, replacement, and annulment of the target memory)

MR would be the clear mechanism of therapeutic change whenever the unlearning trifecta is followed promptly by the lasting disappearance of target memory reactivation and of the unwanted behaviors, affect, cognition, and somatic disturbances it had been producing, as in the case example below in Section 6.2.

The relevance of MR to psychotherapy is very large because the array of clinical symptoms generated by the contents of memory is very large [10, 11, 47] including myriad post-traumatic symptoms, myriad symptoms arising from insecure attachment, the many types of compulsive behavior, addictions, low self-esteem, and many instances of anxiety and depression (such as the depressed woman in the case example below). The aim of clinical translation of the unlearning trifecta is to equip psychotherapists and mental health counselors to fully dispel emotional memory-driven patterns of unwanted behaviors, affect, cognition, and somatic disturbances by fundamentally transforming their mnemonic roots.

6.1 An empirically based therapeutic methodology of transformational change

Full annulment of problematic emotional responses would be far above the standard of effectiveness of the mainstream clinical field, where partial, mild, incremental degrees of symptom reduction have been the norm [13, 14, 15]. Even in "evidence-based… treatments…effect sizes… are small to moderate, gains may not persist, and there is a proportion of patients who derive little or no benefit" [151]. Psychotherapy outcome research literature has routinely defined successful therapeutic outcome in terms of mild reduction of symptoms [152]. In randomized controlled trials (RCTs) and meta-analytic reviews, the measured level of therapeutic improvement for more than 40 years has been a change of about one standard deviation in the mean score on outcome measures, representing typically a 20% to 25% reduction in the measured strength of symptoms [153, 154].

Thus the emergence of an empirically based therapeutic process of full annulment of unwanted emotional responses would be a fundamental reorientation and major advance of the clinical field.

Facilitating experiences of certain types in psychotherapy is not only feasible but already a quite familiar task to many clinicians. In contrast, implementing particular procedures that were used in laboratory studies would be feasible in only a very limited range of clinical situations. In most instances, replicating a procedure would be incompatible with the processing style of the therapy recipient, with the nature of the symptoms and underlying emotional material being addressed, and with the therapist's repertoire of techniques and personal style.

Also, a concrete procedure replication requirement would mislead therapists by implying that sensitive tailoring of the unlearning trifecta's component experiences is not necessary, which would cause treatment failures. The necessity of tailoring the therapeutic process to the individual's unique psychological material and processing style is well established by clinical outcome research [14, 15]. Every person's emotional learning history is unique [47]. Given that the universe of target emotional learnings of persons in psychotherapy is vast if not infinite, obviously the range of concrete procedures that could be used in therapy for creating the needed PE experience and counter-learning is likewise vast if not infinite. Pre-defining any particular procedure(s) for facilitating the unlearning trifecta is therefore strongly counter-indicated.

Successful clinical facilitation would be indicated by observing the same prompt, unambiguous markers of emotional memory annulment as in the studies in Table 1, namely, complete and lasting disappearance of reactivation of the target emotional memory as a subjective experience, and disappearance of all of the target memory's behavioral, somatic, emotional, and cognitive forms of expression. That would be "the induction of amnesia for previously established emotional memory" [29] and would warrant being designated profound change or transformational change [10, 12, 47] to distinguish it from gradual, incremental, partial reduction of symptoms. In every study in Table 1, the researchers regard those markers of change as evidence that MR has occurred, and that logic applies equally to the clinical setting.

An extensive clinical literature already exists documenting (a) the moment-to-moment creation of the three critical experiences in the unlearning trifecta, and (b) verification of prompt appearance of unambiguous markers of transformational change [11, 47, 48, 155] (and for an extensive listing of published case examples indexed by symptom, see https://bit.ly/2tKXdyX; for a list of case examples from diverse systems of psychotherapy, see https://bit.ly/15Z00HQ). Those numerous demonstrations suggest that facilitation of the unlearning trifecta has been found to be not only feasible, but also proves versatile and effective in nullifying a wide range of negative emotional learnings and their symptomatic forms of expression.

Such clinical observations are normally regarded by empirical researchers as anecdotal and therefore are largely disregarded as having little or no empirical evidentiary value. However, it is arguable that this particular body of clinical observations does have empirical evidentiary value consisting of detailed, unambiguous documentation of (a) occurrence of the unlearning trifecta's three critical experiences (or methodology evidence) and (b) prompt, decisive markers of annulment of emotional memory (or unlearning evidence), i.e., symptom cessation with no delay and nothing else done or necessary to prevent recurrence of the target memory's problematic responses. Video recordings of therapy clients reporting abrupt, total cessation of major symptoms in no uncertain terms, with follow-up confirmation of no relapses after many months or years, may warrant being regarded as empirical data fully as valid as skin conductance response measurements in laboratory studies.

The published case examples show both the deliberate and the fortuitous facilitation of the unlearning trifecta. For example, Ecker [47] shows deliberate facilitation for two cases, one dispelling a long-term feeling of terror and intense somatic disturbances (with session videos available for viewing) and the other, a lifelong disposition to a prolonged mood of angry resentment. Deliberate facilitation in another, particularly complex case [48] dispelled several primary features of borderline personality, including extreme compulsive perfectionism, emotional volatility, self-harming behaviors, and suicidality. Sixteen fine-grained cases of deliberate facilitation are presented in [68]. Fortuitous facilitation is shown in a collection of eight cases that ended a wide range of symptoms, with each case involving a different system of psychotherapy [68]. A therapeutic MR protocol that was highly effective for adolescents with PTSD [156] was shown to contain fortuitous facilitation of the unlearning trifecta [47].

The present article is intended to serve as a direct bridge between the extensive body of clinical observations and the laboratory research mapped in Table 1 and Appendix A. It would do that by making it clearly apparent that the same unlearning trifecta that was implemented in successful laboratory studies is present unambiguously in therapy sessions that produce prompt transformational change. Researchers may tend to object that replication of experiences, rather than specific procedures, is inherently too poorly defined and elusive of confirmation to allow for reliable, deliberate, definite replication. However, each of the three requisite experiences in the unlearning trifecta actually is well defined and reliably has salient, distinct, confirmable behavioral features, as shown explicitly in each of the published case examples referenced above, and also in the case example given below in Section 6.2. Furthermore, the equally unambiguous observation of prompt, lasting transformational change is itself a strong indication of fulfillment of the unlearning trifecta, because, as noted previously, only the MR process can produce such change, according to all current scientific knowledge.

Various possible obstacles to clinical translation have been identified by neuroscience researchers who assume that clinical translation would consist of using laboratory procedures [26, 29, 56, 69, 70, 71, 72, 73]. However, Ecker [47] has argued that clinical translation through replication of experiences rather than procedures is not vulnerable to those obstacles. Section 7 below addresses those issues.

A study design described in [13] calls for videos of therapy sessions to be rated by two separate video coding systems and two independent teams of raters, for separately detecting in therapy sessions the component experiences of the unlearning trifecta (methodology evidence) and the markers of transformational change (unlearning evidence), as well as the precise timing between them, across a wide range of symptoms, therapeutic systems and techniques, client/patient characteristics, therapist characteristics, and client-therapist relationship qualities. In that way, the causal role of the unlearning trifecta in producing transformational change could receive empirical support, which would also be empirical evidence of MR as an internal mechanism of therapeutic change.

Such evidence supporting the causality of a specific internal mechanism of change would be an historically significant milestone for the psychotherapy field [13]. The clinical field has inferred many *mediators* of change that are often mislabeled as mechanisms, and has evolved for a century with no empirical identification of an internal mechanism of change, allowing continual fragmentation into competing theoretical systems [157, 158, 159, 160]. Such evidence would also support viewing the unlearning trifecta as a core process shared across psychotherapy systems that produce transformational change regularly. The unlearning trifecta would then be a fundamental part of an empirically supported framework of psychotherapy unification [10, 11]. In addition, transformational change would be de-mystified and could logically become the natural standard of effectiveness for psychotherapy, another significant advance for the clinical field.

6.2 Clinical case example

A short clinical case vignette previously reported [46] will serve to illustrate the clinical facilitation of the unlearning trifecta and the salience and verifiability of the component experiences. It should be borne in mind that each of the three component experiences (target memory reactivation, a concurrent PE experience, and counter-learning) is tailored uniquely in each case, so this particular case does not represent a general template for therapeutic techniques or style. Any clinical case example of the unlearning trifecta is necessarily an example of a choice of particular techniques and style used for facilitating it. Because facilitating it does not depend upon particular techniques or procedures, clinicians are free to utilize their preferred techniques and style to create the three experiences, which may be the single most important practical factor conducive to its appeal and adoption.

A married female therapy recipient in her 50s (there is no identifying information in this account) wanted relief from long-term depression and sexual aversion. The therapeutic work in her tenth session began with a focus on episodic memory. The therapist guided her to revisit imaginally and to subjectively reinhabit and replay her situation at age 18: While living with her parents in a small conservative town, she was unmarried and had become pregnant, and then, while struggling with deciding whether to have an abortion, she had a miscarriage.

She closed her eyes as she began to retrieve the subjective, episodic memory of being in that crisis, and the therapist prompted her to describe, in present tense, the various aspects of her outer life and inner condition. That process evoked the imaginal perception of various potent cues and contexts present in the original experience, which deepened her reactivation of the episodic memory and her subjective immersion in it. Her speech became progressively quieter and slower.

After the miscarriage, she described feeling a heavy, hopeless despair. The therapist softly asked her if there were any words for why she has that heavy feeling of hopeless despair. Several seconds of silence passed, and then she said somberly, “The rest of my life as a woman is ruined. I’ll never marry. I’ll never have children.”

Through the portal of the episodic memory, she was now consciously accessing the mental model of her future that she had construed after that crisis at 18: Permanently stigmatized, no man would ever marry her, so she would never have children, and the rest of her life as a woman was ruined. The therapist immediately understood that this potent construct of meaning, persisting timelessly outside of awareness in implicit, nonverbal, emotional memory, was generating both her depression (the mood of being in hopeless despair and profound grief, but with the specific content of that despair and grief suppressed out of awareness) and her sexual aversion (because enjoying sex is what had ruined her life).

This was the reactivation and retrieval into explicit awareness of the target emotional learning underlying her presenting symptoms. Knowledge of the specific content of the target schema would now guide the therapist in creating the next two experiences of the unlearning trifecta: prediction error and counter-learning. Those crucial next steps can be facilitated by the therapist reliably only on the basis of first bringing the target schema into explicit verbalization, revealing its specific content. (Researchers have shown that conscious awareness of a reactivated memory is not necessary for destabilization, updating, and reconsolidation to occur [92, 126, 161, 162]. However, as a rule, consistent, effective facilitating of the unlearning trifecta in psychotherapy requires both conscious awareness and explicit verbalizing for the reason just given. For fuller discussion of this point, see [46].

The therapist happened to know that in other memory networks she was already in possession of contradictory knowledge that could now be accessed to create both the PE and counter-learning experiences. In order to elicit the contradictory knowledge into concurrent awareness or juxtaposition with the reactivated target schema, the therapist asked her empathetically to say those same words again. She again said, “The rest of my life as a woman is ruined. I’ll never marry. I’ll never have children.” It was clear from her tone and manner that once again the words simply felt true to her, so the therapist gently said, “Say it again.” As she began to say it a third time, immediately her head jerked up, her eyes opened wide, and she said sharply and loudly, “Wait—that’s not true! I did marry! I did have a child!” The PE experience in that moment was definite, and presumably the encoding of her target schema was now rapidly destabilizing. Her boy was now age 14, as the therapist knew. In having her repeat the words of her schema, the therapist was making deliberate use of her brain’s ever-active mismatch detector to find and bring forward the contradictory knowledge.

She was astonished. She sat for some seconds in wordless amazement, with her head fallen back against the top of her chair, gazing up toward the ceiling. Then she quietly uttered, “This is huge,” and described strong sensations of energy streaming throughout her body, down to her toes.

During the remainder of the session, the therapist empathetically kept reviewing the target schema and the contradictory knowing, which repeated their juxtaposition several times, and these repetitions served as counter-learning experiences, as did the repetitions of the juxtaposition that were certainly happening spontaneously in her inner thought process of reorganizing these components of her world of meaning. By the end of the session, the dark, dire meaning that she had attributed to her crisis, and that she had experienced as being completely real and objectively true, had been thoroughly unlearned and nullified. That expectation of her future no longer had any emotional realness and, at least functionally, no longer existed in implicit emotional memory.

After that session, she reported the lifting of both her longstanding depressed mood and her sexual blockage. In a session two weeks later she said, “I’ve initiated a little bit of spontaneous sex, and was interested,” which was a major reversal for her. These changes remained in effect, with no further work needed for maintaining them, during another fourteen months of twenty more therapy sessions addressing other areas.

It is particularly significant to note in the case example above that the specific techniques used for creating the three component experiences were unlike any of the procedures used for creating those experiences in the laboratory studies in Table 1, yet all three experiences were created efficiently and effectively, and produced prompt annulment of the target emotional memory. This again highlights the distinction between experiences and procedures, as well as the freedom of technique and style with which psychotherapists and counselors may facilitate the unlearning trifecta, with ample empathy and emotional attunement to subjective experience and felt meanings.

That case example illustrates also that each of the three critical experiences is well defined, salient, and distinct, allowing reliable facilitation, and that a therapy recipient’s reports of prompt and lasting annulment also are unambiguous. That is the best evidence indicating successful clinical translation of the unlearning trifecta: the combination of unambiguous methodology evidence (distinct occurrence of each of the three experiences), unambiguous unlearning evidence (distinct indications of the disappearance of the target schema and its manifested symptoms), and the proximal temporal sequence between the methodology evidence and the unlearning evidence (the latter appearing next after the former, with no further therapeutic work being applied to the target memory or its unwanted forms of expression). In general, causality is notoriously difficult to prove in the context of psychotherapy [160]. However, the overall evidentiary pattern just described substantially fulfills several causality criteria, namely the combination of a well-defined mediator (activating condition, the unlearning trifecta), a strong and unique effect (annulment, transformational change), and a distinctive timing relationship between mediator and effect that is unique to the proposed causal mechanism [13, 157, 159, 160].

Decisive clinical evidence of long-lasting annulment of emotional memory cannot be explained by any known type of neuroplasticity or learning other than behavioral updating via MR. Regulation (suppression, inhibition) of an existing, acquired emotional response can occur through various types of learning [163, 164], but these are competitive rather than transformational in nature [10, 13, 165], making suppression unstable and allowing re-expression of the suppressed response in a new context or under stress.

In the clinical example above, the needed contradictory knowledge was readily available to emerge via mismatch detection instigated by overt statements of the target schema [166], but that is not always the case. If no contradictory knowledge emerges in that way, the therapist must actively guide a process of either finding or creating it, and a wide range of techniques can be used to accomplish that [166].

Also apparent in that case example is the use of a contradictory knowledge that is non-emotional for disconfirming and nullifying a strongly emotional target memory. That is one of the features listed in Section 4 in summary of the empirical studies reviewed in Appendix A. A heavy feeling of hopeless despair accompanied the target emotional memory, "The rest of my life as a woman is ruined. I'll never marry. I'll never have children." And yet, the contradictory knowledge that juxtaposed with that emotional schema, disconfirming and nullifying it, consisted of her own familiar, non-emotional, factual knowledge that she had married and had a child. The familiar facts that she had married and had a child had no salient emotional quality in themselves for her prior to that juxtaposition, yet they potently disconfirmed and nullified a tenacious, intensely emotional learning formed decades earlier. This may seem counter-intuitive until it is recognized that disconfirmation and unlearning require not emotion, but a direct knowing or perception that decisively contradicts what is known and expected according to the target learning. "The disconfirming new experience or knowledge must be experienced as being unmistakably true and real, but that is not necessarily an emotional experience" [47]. Certainly, that could also be accomplished by a new emotional experience. The point is that a potently disconfirming, contradictory experience does not have to be an emotional experience. "The essence of the target learning is its implicit model of reality (its semantic content), not the emotion that arises from that construal of reality. It is the model that is being disconfirmed and unlearned, not the emotion. …A transformational change in the model immediately produces profound change in the emotion generated" [47].

In the context of psychotherapy, profound change or transformational change resulting from the unlearning and annulment of an implicit mental model or schema have been defined by the following set of markers [10, 47, 68], which are seen in the case described above:

- One or more longstanding symptoms of behavior, state of mind, or somatic disturbance cease to occur.
- The altered state of emotional activation (ego state, self state, emotional schema, core belief) that was underlying the symptom no longer reactivates or feels real.
- Those two changes then persist effortlessly and permanently, with no preventative practices needed for maintaining the symptom-free condition.

7. Potential obstacles to clinical translation identified by neuroscience researchers

A sizable fraction of published empirical studies and review articles on behavioral updating via MR contains statements such as this one: "Previous studies have reported that behavioral extinction training within the reconsolidation window could prevent the recovery of fear memories…, but this line of research has been hampered by partial or full replication failures…, questioning whether the reported effects are stable enough to be translated into complex clinical settings" [131]. Identifying the causes of the replication failures and how those causes may be obstacles to clinical translation has been the topics of numerous review articles [26, 29, 56, 69, 70, 71, 72, 73].

A summary of those discussions has been made by Ecker [47], who listed and examined those authors' five main potential obstacles to clinical translation of emotional memory updating:

- The conditions that create PE experiences may be too exacting for clinicians to create reliably.
- The age and strength of emotional learnings in therapy could make it too difficult to create PE that destabilizes them.
- Therapeutic changes made via reconsolidation might not be durable.
- Episodic memory may be resistant to destabilization and updating.
- Clinicians may not be able to navigate complex memory structures.

Each of those possible obstacles was identified by researchers on the basis of assuming that clinical translation would consist of applying the same concrete procedures that were used in laboratory studies. Each putative obstacle has been reevaluated by Ecker [47] on the alternative basis of assuming that clinical translation would consist of replicating the three experiences in the unlearning trifecta (listed above in Section 6), rather than replicating any study's particular procedures, and found that none of the potential obstacles listed above pose a significant problem with that approach. A full recapitulation of all of those analyses is beyond the scope of this article, so summarized here is the substance of only the first issue:

The concern that PE must be tuned very finely for destabilization to occur, perhaps too finely for clinical success to be reliable, is based on laboratory studies that systematically varied experimental parameters to test a range of degrees of memory mismatch at reactivation [20, 109, 110, 111, 112, 113, 114, 115, 116, 124]. Memory destabilization, inferred from post-reactivation pharmacological abolishing of the acquisition memory, occurred for a very narrow range of mismatch parameters. The analysis of this phenomenology by Ecker [47] points out that "the delicacy of [successful] mismatch observed in laboratory studies is a direct effect of the artificially precise structure of the target learnings created by researchers, not an inherent feature of the reconsolidation process that would necessarily carry over into the clinical setting. …[T]he detailed structure and content of the target learning determine the post-reactivation experiences that will, or will not, mismatch and destabilize the target learning. Thus the finding that Pavlovian target learnings are mismatched and destabilized by post-reactivation procedures that may vary within only a narrow range of parameters is entirely due to the precisely ordered structure of Pavlovian target learnings designed by researchers." (For specific analyses of how the highly structured sequences and timings in laboratory acquisition trainings determine the reactivation conditions that produce PE, see the reviews of individual studies in sections A.1, A.3, A.5, A.7, A.8, A.10, A.11, A.13–A.16, and A.18.) In contrast, in psychotherapy the emotional schemas that are the primary target for annulment have no such precisely ordered structure. Rather, they consist of semantic implicit knowledge of specific contingencies and meanings abstracted from concrete experiences, such as the implicit (nonconscious) knowing that being visibly emotional or energetic will bring immediate harsh shaming, rejection, and disconnection. Another example, from the case example in Section 6.2, is the learned expectation, "The rest of my life as a woman is ruined. I'll never marry. I'll never have children." Ecker [47] states, "mismatching such schemas is a qualitative matter, not a quantitative matter, which eliminates the need for parameter precision in forming mismatches. For example, no careful adjustment is needed on duration of reactivation of a schema that is being mismatched by a contradictory knowing, whereas for mismatching a Pavlovian conditioning memory [part of which is the well-defined timing structure of the acquisition training], duration of reactivation is a critically sensitive parameter." In short, "Emotional learnings retrieved in therapy are found to be wide targets for a destabilizing mismatch, not narrow targets" (p. 52).

8. Critique of proposed frameworks of therapeutic memory reconsolidation: An appraisal of fidelity to empirical findings

Section 4.1 lists the core findings of the first 20 empirical, statistically strong MR studies demonstrating emotional memory unlearning and annulment in humans. An important use of that set of empirical findings is to constrain emerging accounts of the psychotherapeutic functioning of MR to assure scientific fidelity.

Historically, the psychotherapy field has had little engagement in the normal process of scientific development in which the emergence of confirmed empirical knowledge necessarily imposes constraints on subsequent hypothesizing and modeling [158]. Psychiatrists, psychologists, and psychotherapists have long enjoyed a free hand in inventing theoretical frameworks of mental functioning and processes of therapeutic change, allowing the fragmentation of the clinical field into hundreds of systems of psychotherapy, each with its own concepts, methodology, and techniques [167] Some eminent voices have been calling attention to the crisis of the clinical field drifting without an empirical foundation, without, in fact, empirical knowledge of an internal mechanism of change [13, 157, 158, 159, 160].

In that context, the emergence of MR neuroscience and, in particular, laboratory demonstrations of utilizing MR for emotional memory annulment in humans, are viewed within the clinical field as having enormous potential to advance the clinical field on several fronts [10, 11, 13, 47, 168]. Virtually all new psychotherapy texts, workshops and trainings now include material purporting to show how the presented therapeutic approach embodies and utilizes MR. Alongside such embracing of MR on behalf of particular systems of therapy, several general accounts of therapeutic MR have also been published by psychiatrists, psychologists and psychotherapists.

The arrival of empirical constraints and correctives for those accounts would be an unfamiliar, major change in the landscape of the clinical field, but that moment appears to have arrived due to the studies listed in Table 1 and their consensus findings, listed in Section 4.1, reaching a critical mass. The clinical field's optimal use of MR research findings depends on the accurate transmission and understanding of this empirical knowledge. This section is therefore an initial attempt to enter this new territory by reviewing the several proposed general accounts of therapeutic MR in order to examine their degree of consistency with the empirical findings listed in Section 4.1. It will be suggested below that the clinical field's longstanding license to follow theoretical biases has already been shaping much of how therapeutic MR has been interpreted and presented, at the expense of scientific accuracy, and thus clinical effectiveness.

Four general accounts of the therapeutic occurrence and use of MR have been formulated, by Ecker and colleagues [10, 13, 46, 47, 68, 169, 170], by Lane, Ryan, Nadel and Greenberg [168], by Stevens [171], and by Welling [172].

8.1 Ecker and colleagues

The therapeutic MR framework of Ecker and colleagues, referenced just above, has emphasized its close, explicit, extensive correspondences with the empirical findings delineated in Section 4.1 and is identical to the clinical translation framework mapped out in Section 6.1. That is, it consists of facilitating the unlearning trifecta in psychotherapy sessions and verifying unlearning and annulment by actively confirming the appearance of the markers of transformational change defined above at the end of Section 6. All of those features are apparent in the case example in Section 6.2, which was reported originally in [46] The freedom of clinicians to induce the component experiences of the unlearning trifecta using any suitable experiential techniques is strongly emphasized by Ecker and colleagues.

The MR framework of Ecker et al. also maps out the preparatory steps necessary in psychotherapy for maximum consistency and efficiency of facilitating the unlearning trifecta across diverse cases, given that at the start of psychotherapy, the therapist has no knowledge of the target emotional learning(s) generating the problem patterns. In contrast, laboratory MR researchers know, from the start of a study, the detailed content of the target learning and, therefore, how to create PE and counter-learning experiences. That handicap is addressed in the methodology of Ecker and colleagues by first identifying the specific features of the problem patterns, then revealing the emotional learnings or schemas underlying and driving those problem patterns through experiential processes, and then, guided by the specific content of that revealed material, finding or creating contradictory knowledge or contradictory experience. Those ingredients may require several sessions to assemble, but then in minutes they are then combined to carry out the unlearning trifecta. Those preparatory steps too are illustrated by the case example in Section 6.2.

Ecker and colleagues also indicate how, in addition to the prospect of enhancing the effectiveness of individual psychotherapists and counselors, several other significant advances of the clinical field may be driven by the empirical findings summarized in Section 4.1 [10, 11, 13, 47, 68]:

- identification of an internal mechanism of lasting change
- transtheoretical unification of the severely fragmented psychotherapy field
- new explanation of decades of uniform results from randomized control trials (“Dodo bird verdict”) and resolution of the debate over the roles of common and specific therapeutic factors
- refinement of the “corrective experience” framework of psychotherapy
- clarification of the therapeutic functions of the client-therapist relationship
- clarification of the debate over the therapeutic prevalence of insecure attachment

Disagreement with the therapeutic MR framework of Ecker and colleagues was expressed by neuroscientist Alberini [173], on the basis of “findings that the reconsolidation of complex episodic memories is a temporally limited process. How, given this temporal limit, can reconsolidation explain the updating of old memories? I must disagree, then, with the idea that reconsolidation fully explain [sic] the process of change in psychoanalysis and psychotherapy (Lane et al., 2015; Ecker, Ticic, & Hulley 2012).”

However, Ecker and colleagues do not maintain that MR can “fully explain” all psychotherapeutic change processes, so Alberini’s critique of their MR framework is specious. Rather, they maintain that MR alone can explain clinical observations of lasting, transformational change, i.e., the full annulment of a particular emotional schema, or mental model, and the behaviors, emotions, cognitions, and/or somatic disturbances it was generating. Furthermore, the account by Alberini [173] regards episodic memory as the focus of therapeutic MR, whereas in the framework of Ecker and colleagues, the MR process is applied to emotional schemas in semantic memory, which is what makes transformational change possible even as episodic memories remain largely intact.

8.2 Lane et al. [168], Stevens [171], and Welling [172]

As in the methodology of MR proposed for clinical translation by Ecker and colleagues and described in the previous section, Lane et al. [168], Stevens [171], and Welling [172] have also proposed that MR is the fundamental mechanism responsible for lasting change produced in any system of psychotherapy and that identifying the same core process of change operating in diverse therapy systems can be a unifying framework. However, those three accounts characterize the MR mechanism, and the therapeutic process that engages it, very differently from the unlearning trifecta specified by Ecker and colleagues. Examining those differences through the lens of MR empirical research findings occupies the remainder of this section. It warrants emphasizing that these three articles contain a great deal of valuable description of therapeutic process, therapeutic skill, and the role of memory in symptom production, as distinct from their accounts of how MR functions therapeutically to produce change. Only the latter area is the focus of the critiques presented below.

Lane et al. [168], Stevens [171], and Welling [172] offer accounts that are divergent from or contradicted by MR research findings in these major ways:

- They define therapeutic MR as fundamentally occurring through emotion and as requiring the therapy recipient to have a new, healthy emotional experience in the problem context.
- They give no recognition of the brain’s requirement of a prediction error experience for activating the MR process or of the extensive body of research on the PE requirement.
- They give no recognition that extensive MR research has identified a specific process (the unlearning trifecta) that

produces prompt, profound unlearning and annulment of a target emotional schema and its manifestations, and they characterize a different, less effective process.

Those divergences from empirical findings are explained in the subsections that follow. They appear to be the result of constructing accounts of MR with little reference to the MR research literature.

The unlearning trifecta is a specific process of unlearning and annulment identified empirically in neuroscientists' laboratory studies listed in Table 1 and reviewed in Appendix A. However, the accounts by Lane et al. [168], Stevens [171], and Welling [172] contain virtually no recognition of or references to the numerous MR studies that have documented the unlearning trifecta and the PE requirement (https://bit.ly/2b8IbJH). Instead, they appear to interpret the therapeutic process of MR in an ad hoc manner according to familiar psychotherapy concepts and assumptions derived from clinical observations, clinical outcome studies, and models of psychotherapy. Maintaining that interpretive license to define therapeutic MR in theoretically convenient ways would forfeit the current, historic opportunity to understand the mechanism of lasting therapeutic change according to fundamental empirical knowledge that emerged independently of the clinical phenomena that it explains.

If MR research had revealed only that new experiences can revise the contents of memory, and nothing more, then psychotherapists and psychologists would continue to have license to draw upon their own field's knowledge and observations to propose models of how MR operates and what has to happen in therapy sessions to recruit the MR process effectively. Indeed, psychotherapists and psychologists have been free to exercise such interpretive license for a century in the absence of an empirically characterized internal mechanism of change. However, MR research by neuroscientists has revealed far more than the fact that the contents of memory are revisable. It has revealed the details of both the neurobiological mechanism and the operational, behavioral process required by the brain for recruiting MR to nullify an emotional learning, as shown in each of the 20 studies reviewed in Appendix A and summarized in Section 4.1. As always in science, this emergence of confirmed empirical knowledge now necessarily imposes constraints on MR-related hypothesizing and modeling, significantly limiting the interpretive free hand that psychotherapists and psychologists have long enjoyed.

The following subsections are an initial exercise in this new domain where the validity of proposed therapeutic methodologies and principles can be evaluated using strongly established empirical criteria.

8.2.1 Emotion-centric accounts of MR are contradicted by MR research

Lane et al. [168], Stevens [171], and Welling [172] depict the therapeutic action of MR as being an emotion-centered process. They define the target of change as problematic emotion and they describe the MR process in therapy as consisting of arousal of a client's problematic emotional state followed soon by the creation of a strong, healthy emotional experience in the same context.

Welling [172] states, "A maladaptive emotion can only be undone by a sufficiently strong emotion…." Stevens [171], states, "a new experience should be offered through which affect reconsolidation can occur, transforming the negative emotion into a positive emotion." (Welling and Stevens address only problems of affect and give no consideration to how MR would be utilized for presenting problems that are primarily behavioral, cognitive, or somatic.) Lane et al. [168] advance the same view of therapeutic MR in stating (p. 3), "we propose that change occurs by activating old memories and their associated emotions, and introducing new emotional experiences in therapy enabling new emotional elements to be incorporated into that memory trace via reconsolidation" (italics added); and "This model highlighting the importance of new emotional experiences...." (p. 16). That emotion-centric therapeutic strategy and methodology has been articulated most succinctly by coauthor Greenberg in the motto "changing emotion with emotion" [176, 189], which was stated three times by Lane et al. [168].

Emotion and emotional processes certainly are, and should be, major aspects of psychotherapy sessions, as a rule, as in the case example in Section 6.2. That is not at issue here. The issue is that extensive MR empirical findings contradict any account asserting that the MR mechanism in itself inherently entails, involves, or requires emotion or emotional arousal. No such fundamental role of emotion in the MR mechanism or process is reported in the MR research literature (which, as noted, is referenced very little by these authors). Therefore, any emotion-centric account of how MR works appears to be a mistaken theoretical bias of the author(s). MR research has shown that both the initial destabilization of a target learning and its subsequent updating during the reconsolidation window can occur with no involvement of emotion, as reviewed in [46, 169]. In many laboratory studies of MR, the target learning involved no emotion whatsoever, such as memory of a spatial arrangement of emotionally neutral objects [102] or a particular series of finger movements [174], and these target learnings were shown to destabilize, update and reconsolidate also without any involvement of emotion, through the same set of experiences as when the target learning is emotionally laden, such as Pavlovian fear conditioning.

The MR process does require the reactivation of the target learning. In psychotherapy, the relevant target learning has a significant emotional component or accompaniment, as a rule, and the reactivation required for recruiting MR tends to make the target learning's emotional component strongly salient, as seen in the case example in Section 6.2. For that reason, emotional arousal nearly always accompanies the MR process in psychotherapy, yet is neither a fundamental feature of the MR mechanism nor what the MR process of change acts upon, contrary to the assertions of Lane et al. [168], Stevens [171], and Welling [172]. For transformational change, it is a symptom-generating mental model or schema that undergoes the MR updating, unlearning, and annulment process, not an emotion. Adequate reactivation of that target schema necessarily entails arousal of the schema's emotional accompaniment, but it is the semantic content of the schema, not the emotion generated by the schema, that is subjected to disconfirmation and unlearning via MR. Unwanted emotion disappears as an immediate derivative result of unlearning and nullifying the problematic mental model that had been generating it, as is apparent in the case example in Section 6.2.

In laboratory MR studies that have nullified Pavlovian fear, the target of change was the learned schema consisting of the expectation that, for example, an audio tone will be followed by the pain of an electric shock. That expectation generates the emotion of fear and the measurable physiological and behavioral expressions of fear, so the target of disconfirmation and change in MR behavioral updating studies is that learned expectation, not the emotion of fear that it generates (which is why some researchers prefer the term "threat memory" rather than "fear memory" to describe the object of their study). With that expectation unlearned and nullified, the fear generated by that expectation no longer is produced.

Lane et al. [168], Stevens [171], and Welling [172] make the ad hoc assertion that eliminating a problematic emotion via MR requires a different, healthy emotional experience in the same context. That too is refuted by MR research findings, and also by numerous clinical accounts, including the case example above in Section 6.2, showing that a disconfirmational experience does not have to be emotional to nullify a strongly emotional target learning (46, 47, 169). The target learning in Section 6.2 is an implicit emotional schema that became conscious and voiced as, "The rest of my life as a woman is ruined. I'll never marry. I'll never have children." Disconfirmation and unlearning of that schema came from its juxtaposition with her own familiar, non-emotional, factual knowledge that she had in fact married and had a child. A non-emotional contradictory knowledge also was used in two case examples [47] in which lifelong anger reactions and severe post-traumatic emotional and somatic symptoms were ended promptly by profound unlearning of the symptoms' underlying implicit schema (mental model). Emphasizing the importance of this phenomenology for clinicians, Ecker [170] wrote, "If therapists believe that only a new experience that is distinctly emotional in itself can serve to disconfirm and rewrite target emotional learnings, they would be precluded from utilizing a major class of options for facilitating effective juxtaposition experiences" (i.e., concurrent experience of the target learning and contradictory knowledge, fulfilling the unlearning trifecta).

Thus, although emotion-centric accounts of MR are well aligned with the recent period's conventional, widely held concepts in the clinical field, such accounts are a misrepresentation of the MR mechanism of unlearning and annulment, according to both empirical MR research and clinical observations of the annulment process producing transformational change.

The emotion-centrism of Stevens [171] is the most extreme of the three emotion-centric accounts because it proposes explicitly that the term affect reconsolidation should be used instead of memory reconsolidation in the clinical context. How Stevens uses the term reconsolidation is exemplified by the statement, "it is unlikely that one can reconsolidate affect if the affect cannot be regulated to a tolerable level of arousal…" (p. 4). The actual neuroscientific meaning of reconsolidation is completely lost in Stevens' account. In neuroscience that term denotes the biochemical and neurophysiological transition of a memory's neural encoding from a stable state in long-term, consolidated memory to an unstable, deconsolidated state of susceptibility to modification, and then back into a stable, reconsolidated state (with or without having been modified while destabilized). In Stevens' use, however, reconsolidation is merely a synonym for "change" that appears to reference neuroscience but actually is devoid of neuroscientific meaning. Neuroscientists do not and would not write "reconsolidate affect" because what undergoes the process of reconsolidation is a memory encoding, not an affect. For that reason, memory reconsolidation is the accurate term and affect reconsolidation embodies an incorrect conceptualization that would prevent accurate understanding of the MR mechanism of therapeutic change.

As noted, a symptom-generating schema must be reactivated in psychotherapy in order for it to go through the MR process, and that schema nearly always has a significant emotional accompaniment, so the required reactivation of the schema entails arousal of that emotion. That is how the MR mechanism explains the strong positive correlation between therapeutic change and emotional arousal, consistently documented by psychotherapy outcome researchers, such as in [175], and observed by clinicians. If the emotion associated with a schema has itself been intolerable and therefore chronically avoided and suppressed, the schema itself is all the more strongly kept out of awareness. Conversely, facilitating tolerable experiences of the previously avoided emotion allows the emotion's underlying schema to also come into awareness, which in turn allows the schema to be subjected to disconfirmation and unlearning in the MR process of transformational change.

The schema or mental model is almost always implicit initially, that is, outside of awareness, but the problematic emotion or mood that it generates is consciously felt, is apparent to the therapist, and tends to be regarded as the problem and the target of change, when actually the existence of the underlying schema is the problem and the target of the MR process of transformational change. Of course, one of the therapist's tasks is preventing the level of emotional arousal from becoming too intense for the client to engage usefully in the therapeutic process. Managing the emotional process moment to moment requires considerable skill and can require much of the time and attention available in a session. That necessary preoccupation with emotional process is probably another factor that (mis)leads some clinicians and researchers to an emotion-centric account of the process of lasting change via MR.

A published case of Emotion-Focused Therapy [176], conducted in adherence to "changing emotion with emotion," is examined in a moment-to-moment manner by Ecker et al. [68], showing that the client's lasting change of emotion occurred immediately after the component experiences of the unlearning trifecta were created and had disconfirmed and nullified a mental model that had been outside of awareness. The component experiences of the unlearning trifecta were embedded in the therapy work, were not identified during the sessions or in the published case study, and, importantly, are not identified when the process of change is described as "changing emotion with emotion". Regarding that case analysis, Ecker [47] explained, "By combining emotional reactivation of the target learning with a very different emotional experience of the original situation, a disconfirmation of model occurs implicitly, even though the attention of therapist and client may be focused on derivative emotion." Thus, therapy sessions conducted according to the methodology of "changing emotion with emotion" can and often do produce transformational change. That too is not at issue. The issue is the clinical field's formulation of a scientifically accurate and complete account of the actual process and phenomenology of transformational change via MR.

8.2.2 Any adequate general account of therapeutic MR must recognize the PE requirement

The accounts of therapeutic MR provided by Lane et al. [168], Stevens [171], and Welling [172] do not recognize or mention the brain's requirement of a prediction error (PE) experience for destabilizing a target learning, initiating the MR process. That requirement is one of the major findings of MR research, as described above in Sections 3.1 and 4.1, and it is an indispensable part of any adequate account of MR. The necessary PE experience is an explicit component of the unlearning trifecta. At least thirty separate studies since 2004 (listed online at https://bit.ly/2b8IbJH) have shown that memory reactivation alone, without PE, does not induce target learning destabilization. However, Lane et al. [168] assert the long disproven view that each memory reactivation alone is destabilizing: "…every time a memory is retrieved, the underlying memory trace once again enters into a fragile and labile state" (p. 12); and "We believe that recalling old memories with some details of the original context puts that memory into a labile state…" (p. 51). Likewise, Stevens [171] propagates the same disproven view that in MR research, "if a memory became reactivated, it also became labile…." (p. 2).

8.2.3 Any adequate general account of therapeutic MR must recognize its capability for prompt transformational change

In each of the 20 empirical studies showing annulment of an acquired emotional response (listed in Table 1 and reviewed in Appendix A), the absence of the target response is first apparent immediately upon completion of the unlearning trifecta and then is confirmed at 24 h when the response cannot be re-evoked by presentations of either the CS+ (spontaneous recovery test) or the US (reinstatement test). As noted in Section 4.1, long-term persistence of annulment follows with nothing else done to prevent the response from recurring.

In clinical facilitation of the unlearning trifecta, the same promptness of annulment is observed, as illustrated by the case example in Section 6.2 and as shown in numerous other published case examples with moment-to-moment documentation of process [11, 47, 68, 170]. Such profound change is radically different from the gradual, incremental process of change that has conventionally been assumed and expected throughout the field of psychotherapy, as described in Section 6.1 above and in [13].

Lane et al. [168] adhere to that conventional assumption by defining the therapeutic MR process as requiring that the change-inducing learning experience has to be followed by a final step of "practicing a new way of behaving and experiencing the world in a variety of contexts" (p. 1). That is restated later in their article as the necessity of "repeatedly experiencing and 'working through' the emotional consequences of new learning in a variety of contexts" (p. 49). As defined by Lane et al., therapeutic change via MR requires an extended process of "working through…in a variety of contexts".

The "working through" concept of therapeutic change is one of the oldest and most familiar conceptualizations in the history of psychotherapy, originating with Freud and maintaining currency in psychoanalytic and psychodynamic systems to this day. Many clinicians and clinical theorists take it to be an axiomatic truth of the process of therapeutic change. Certainly such an extended process is necessary, as a rule, when change is being facilitated through competitive learning, i.e., the learning and building up of preferred new behaviors, beliefs, attitudes, and/or states of mind intended to happen instead of, and preventative of, the problem pattern, the hidden mnemonic basis of which remains intact.

However, such an extended process of competitive change is fundamentally different from the promptness and completeness of transformational change that the unlearning trifecta produces, as seen in each study reviewed in Appendix A, in the case example in Section 6.2, and in numerous published case examples [11, 47, 68, 170]. This qualitative and striking difference between the extended final step of the process of change according to Lane et al. [168] and the prompt, profound change produced by the unlearning trifecta seems attributable to the absence of consideration by Lane et al. of the extensive MR behavioral updating research, listed in Table 1 and reviewed in Appendix A, that has established the unlearning and annulment phenomenon and identified the unlearning trifecta.

In psychotherapy, all three component experiences in the unlearning trifecta must happen in one therapy session, and when facilitated successfully, the markers of transformational change promptly begin to appear either later in the same session or in daily life after that session, with no need for the type of extended process called for by Lane et al. [168] The "working through…in a variety of contexts" that Lane et al. prescribe as the final step of the therapeutic MR process seems to be an ad hoc interpretation of MR through the lens of conventional assumptions of psychodynamic psychotherapy. It is a significant element of divergence from MR empirical research findings. Lane et al. give no indication of recognizing the existence of the unlearning trifecta, the extensive research that has identified it, or the phenomenon of transformational change, i.e., the prompt unlearning and annulment of emotional memory. They never describe or call for the "disconfirmation" of a problematic emotional learning.

Rather, the therapeutic process of MR delineated by Lane et al. consists of the conjoining (integration, linking) of desired new memory contents with the existing problematic contents. Such conjoining of old and new memory contents has been reported in human MR studies [146, 147], as noted earlier. That is a qualitatively different and weaker type of memory modification via MR than the unlearning and nullification effect that results from the unlearning trifecta through a highly specific disconfirmation of the reactivated target schema's particular model of reality by a contradictory knowing. Conjoining, as described by Lane et al., involves only coactivation of the negative problematic response with a positive experience. That is not a fundamental, ontological disconfirmation. The result of such conjoining is a memory interference effect that weakens but does not profoundly eliminate the response of the target memory.

The limited therapeutic goal of conjoining is clearly apparent in the writings of psychologists who acquire and use the therapeutic MR model of Lane et al. [168]. For example, [177] state, "According to the…model of Lane et al. (2015), …the changes …would occur by activating old memories and their associated emotions, and introducing new salutary emotional experiences…. This process would enable new emotional elements and a more flexible self-view to be incorporated into autobiographical memory traces via the reconsolidation process."

Importantly, only transformational change (total, lasting disappearance of symptoms and underlying schemas) may be reliably inferred to be due to MR, because only MR can (endogenously) produce such lasting annulment of emotional memory, according to all current scientific knowledge, as discussed in Section 2.1. Lesser degrees of change, such as is produced by conjoining, can have other possible causes, confounding the identification of the operative mechanism. Lesser degrees of change are what Lane et al. [168] imply by maintaining that an extended process of working through is necessary even when MR has been recruited. Thus, the proposed MR process and methodology of Lane et al. are inherently ambiguous as to whether MR is actually operative. In psychotherapy as in the laboratory human studies of annulment in Appendix A, only full, lasting annulment may be regarded as strong evidence that MR has been the operative mechanism of change. However, Lane et al. never discuss emotional memory annulment or the transformational change that it produces.

Facilitation of the unlearning trifecta in psychotherapy occurs within a web of various internal and external complexities, which sometimes require ecological adjustments of the therapeutic process, but these adjustments are not the working-through prescribed by Lane et al. [168] For completeness, the adjustments most commonly found necessary are noted here:

- A prompt transformational change achieved in one context without any working through, such as the workplace context, might not generalize to some other context, such as family relationships, so the unlearning trifecta has to be facilitated in the other context, where transformational change then also occurs without any working through.
- Any symptom-generating emotional schema targeted by the unlearning trifecta has been part of how the individual organizes and responds to the world, so its unlearning and annulment portend adjustments, which in some cases will be uncomfortable. If the individual's implicit knowledge system foresees more discomfort than it deems tolerable, disconfirmation and unlearning is blocked and the target schema continues to feel real and remains in force, despite skillful facilitation of the unlearning trifecta. That unique phenomenology of resistance to schema nullification has been described in [47, 68]. It requires a process of consciously identifying the anticipated adjustment distress

and finding how to render it workable. When that is accomplished, repeating the unlearning trifecta succeeds, the schema is promptly nullified, and no working through is needed.

- In cases where the target schema was learned in traumatic experiences of extreme distress or endangerment, the initial process of bringing the schema into awareness, revealing it, must be done sufficiently gradually, in "small enough steps" [47, 166] to feel workable, avoiding retraumatization, high anxiety, dissociation, or freezing. This gradual process facilitates the initial, conscious accessing of the symptom-generating material, before any process of change has occurred, so it does not constitute the "working through" of a shift that has already occurred.

In summary, this section has applied this article's MR research review to evaluate the scientific fidelity of the four general accounts of therapeutic MR proposed in the peer-reviewed literature. It appears that the accounts of therapeutic MR by Lane et al. [168], Stevens [171], and Welling [172] diverge in several major ways from well-established empirical MR research findings. Recognition of these significant departures from scientific validity should limit the degree to which these accounts of MR will be considered by the clinical field going forward. The full therapeutic potential of MR is very great, but it cannot be realized by adhering to the templates these authors have defined.

9. Conclusion

Re-examining laboratory studies that have reported MR behavioral unlearning of emotional memory in humans, as well as studies that reported non-replication of that effect, has produced these main conclusions:

- It is feasible, based on the preponderance of empirical evidence indicating the necessity of prediction error (PE) for destabilizing memory encoding, to define provisionally a framework of criteria for principled identification of specific PE experiences occurring during experimental procedures, including recognition of the dynamical relationship between procedure mismatches, internal PE experiences, and construal of latent causes. This account of PE phenomenology is considerably more complete than was previously considered in the MR research literature.
- In this trial PE evaluation framework, defining mismatch as a feature of external procedures and PE as an internal experience proves clarifying for analyzing cause and effect in experimental MR procedures.
- Using this PE framework to evaluate human updating replication failures has led to the testable conclusions that (a) an absence of PE experience is the probable cause of all replication failures and (b) the governing boundary conditions are the thresholds where procedural mismatch becomes too small or too large to create a PE experience, given the totality of target memory content at reactivation.
- The most probable cause of absence of PE in most cases of non-replication is found by this framework to be a too-large mismatch of target memory contents by reactivation or post-reactivation conditions, inducing a change of latent cause, i.e., a contextual disconnection from the target memory.
- The most common excessive mismatch is found by this framework to be caused by a previously overlooked element of experimental procedure, the communications made to subjects in transitioning from memory reactivation to a 10-min pause before extinction training begins, causing subjects to regard the 10-min period and its contents (video, magazines, or resting) as a "break" that is not part of the study. This conclusion could be tested by repeating the study with subjects told at the start of the memory manipulation day, "Each of the different phases that will happen today is part of the study" and then zero communication to the subject during the actual transition from memory reactivation to the 10-min interval, with all equipment remaining connected to the subject in the "on" setting. That change of procedure is predicted to result in successful updating and annulment of the target memory for studies that first resulted in replication failure.
- Mismatch that was too small (insufficiently salient) to create a PE experience was probable in one non-replication study.
- The PE analysis and conclusions are empirically testable because for every study that failed to replicate the updating and annulment effect, a specific change of procedure is defined in Appendix B for each such study that should create a PE experience and thereby result in replication.
- Successful updating studies used a wide range of external procedures, yet in each case, the procedure created in subjects the same invariant set of three internal experiences: the reactivated state of the target memory, a prediction error experience during memory reactivation, and a counter-learning experience contradicting the knowledge and expectations in the target memory.
- That set of three experiences, which has been termed here the unlearning trifecta, can be understood as being the fundamental mediator that drives the unlearning and annulment of emotional memory through the mechanism of memory reconsolidation.
- Laboratory empirical demonstrations of emotional memory updating have gone beyond fear memory to more complex emotional learning acquired through processes more complex than Pavlovian conditioning, and beyond the retrieval-extinction protocol to a more general unlearning trifecta methodology.

- The unlearning trifecta is defined independently of any laboratory or therapeutic procedures that may be used to create the three necessary experiences.
- Clinical translation has been defined as replication in psychotherapy sessions of those three experiences using any suitable techniques, not replication of any particular procedure used in laboratory studies.
- Psychotherapy sessions containing unambiguous replication of the unlearning trifecta can validly be regarded as a direct translation of MR research findings, with verification consisting of a psychotherapy recipient's indications of promptly resulting transformational change, i.e., lasting abolishment of an emotional response/schema and the symptom(s) it was generating.
- As shown in a clinical case example, the component experiences in the unlearning trifecta are typically distinct and salient in therapy, and therefore readily detectable and verifiable, as are the resulting, unique markers of change. This indicates the viability of conducting qualitative empirical studies of the clinical occurrence and causal role of the unlearning trifecta that could even more firmly identify memory reconsolidation as the internal mechanism of transformational therapeutic change.
- Comparing the empirical findings of successful reconsolidation updating studies with proposed frameworks of therapeutically applied memory reconsolidation has revealed significant departures from scientific fidelity in three of those accounts.

The clinical field potentially could benefit from various sizable advances due to the infusion of knowledge from MR research, as summarized in Section 6.1. Clinical translation defined as replication of the unlearning trifecta in therapy sessions would be a seamless joining of neuroscience, which embodies an ethos of technical procedures for acquiring impersonal, objective knowledge of brain phenomena, and psychotherapy, which embodies an ethos of sensitive, empathetic attunement for acquiring knowledge of persons' subjective experiences of states of mind. That synthesis would be a remarkable achievement.

This review's analysis was limited to human studies of behavioral updating but, as noted, the same quandary of non-replication exists for human studies of the pharmacological annulment of emotional memory using propranolol. Applying this review's PE evaluation framework to that set of studies would be an important further test of this framework's explanatory capability, and is therefore a desirable future extension of this exploration.

Testable predictions have been made in each of the reviews of non-replication studies in Appendix B. If the predictions were to be confirmed and this review's analyses were to acquire empirical support, the implication would be that the meaning of the replication failures (Table 2) is not that MR behavioral updating has a limited range of clinical applicability, as has been widely suggested, but rather that no limits of behavioral updating have yet been found, because non-replications were due to specific, avoidable experimental conditions that unnecessarily caused procedural reactivation mismatches to be too large or too small to create a PE experience. The updating and annulment effect would then be recognized as being reliable and sturdy rather than elusive and fragile. In that case, the outlook for effective, versatile clinical application would be bright, as the existing body of clinical case documentations has already begun to indicate.

Acknowledgments

The author expresses gratitude to Robin Ticic for discussions that helped to develop Figure 1, to Dr. Melissa M. Reading and Robin Ticic for numerous finalizing improvements of the manuscript, and to neuroscience researchers Dr. Alejandro Delorenzi, Dr. Javiera Oyarzún, Dr. Timothy Jarome, and Dr. Emiliano Merlo for past consultations that contributed to the formation of the current framework.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The author is the unpaid co-director of a clinical training institute that provides training to licensed mental health professionals in the clinical methodology described in this article. The author receives revenues from online sales of webinars, videos, books, and articles that teach and demonstrate this clinical methodology.

References

1. Riccio DC, Millin PM, Bogart AR. Reconsolidation: A brief history, a retrieval view, and some recent issues. Learn. Mem 13 (2006): 536-544.
2. Nader K, Schafe GE, LeDoux JE. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature 406 (2000): 722-726.
3. Pedreira ME, Pérez-Cuesta LM, Maldonado, H. Reactivation and reconsolidation of long-term memory in the crab Chasmagnathus: Protein synthesis requirement and mediation by NMDA-type glutamatergic receptors. J. Neurosci 22 (2002): 8305-8311.
4. Pedreira ME, Maldonado H. Protein synthesis subserves reconsolidation or extinction depending on reminder duration. Neuron 38 (2003): 863-869.

5. Monfils MH, Cowansage KK, Klann E, et al., Extinction-reconsolidation boundaries: Key to persistent attenuation of fear memories. Science 324 (2009): 951-955

6. Schiller D, Monfils MH, Raio CM, et al., Preventing the return of fear in humans using reconsolidation update mechanisms. Nature 463 (2010): 49-53.

7. Clem RL, Schiller D. New learning and unlearning: Strangers or accomplices in threat memory attenuation? Trends Neurosci 39 (2016): 340-351.

8. Fernández RS, Bavassi L, Forcato C, et al., The dynamic nature of the reconsolidation process and its boundary conditions: Evidence based on human tests. Neurobiol. Learn. Mem 130 (2016a): 202-212.

9. Nader K. Reconsolidation and the dynamic nature of memory. Cold Spring Harb. Perspect. Biol 7 (2015): a021782.

10. Ecker B. A proposal for the unification of psychotherapeutic action understood as memory modification processes. J. Psych. Integ 34 (2024): 291-314.

11. Ecker B, Bridges SK. How the science of memory reconsolidation advances the effectiveness and unification of psychotherapy. Clin. Soc. Work J 48 (2020): 287-300.

12. Ecker B, Toomey B. Depotentiation of symptom-producing implicit memory in coherence therapy. J. Constr. Psychol 21 (2008): 87-150.

13. Ecker B, Vaz A. Memory reconsolidation and the crisis of mechanism in psychotherapy. New Ideas Psychol 66 (2022): 100945.

14. Norcross JC, Lambert MJ. Psychotherapy relationships that work: Volume 1: Evidence-based therapist contributions. Oxford University Press, Oxford, UK (2019).

15. Norcross JC, Wampold BE. A new therapy for each patient: Evidence-based relationships and responsiveness. J. Clin. Psychol 74 (2018): 1889–1906.

16. Bouton ME. Context and behavioral processes in extinction. Learn. Mem 11 (2004): 485-494.

17. Vervliet B, Craske MG, Hermans D. Fear extinction and relapse: State of the art. Annu. Rev. Clin. Psychol 9 (2013): 215-248.

18. Duvarci S, Mamou CS, Nader K. Extinction is not a sufficient condition to prevent fear memories from undergoing reconsolidation in the basolateral amygdala. Eur. J. Neurosci 24 (2006): 249-260.

19. Duvarci S, Nader K. Characterization of fear memory reconsolidation. J. Neurosci 24 (2004): 9269-9275.

20. Merlo E, Milton AL, Goozee ZY, et al. Reconsolidation and extinction are dissociable and mutually exclusive processes: Behavioral and molecular evidence. J. Neurosci 34 (2014): 2422-2431.

21. Merlo SA, Santos MJ, Pedreira ME. Identification of a novel retrieval-dependent memory process in the crab Neohelice granulata. Neuroscience 448 (2020): 149-159.

22. Reichelt AC, Lee JLC. Memory reconsolidation in aversive and appetitive settings. Front. Behav. Neurosci 7 (2013): 1-18.

23. Elsey JWB, Van Ast VA, Kindt M. Human memory reconsolidation: A guiding framework and critical review of the evidence. Psychol. Bull 144 (2018): 797-848.

24. Clem RL, Schiller D. New learning and unlearning: Strangers or accomplices in threat memory attenuation? Trends Neurosci 39 (2016): 340-351.

25. Kredlow MA, Unger LD, Otto MW. Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. Psychol. Bull 142 (2016): 314-336.

26. Lee JLC, Nader K, Schiller D. An update on memory reconsolidation updating. Trends Cogn Sci 21 (2017): 531-545.

27. Meir Drexler S, Wolf OT. Behavioral disruption of memory reconsolidation: From bench to bedside and back again. Behav. Neurosci 132 (2018a): 13-22.

28. Zuccolo PF, Hunziker MHL. A review of boundary conditions and variables involved in the prevention of return of fear after post-retrieval extinction. Behav. Processes 162 (2019): 39-54.

29. Beckers T, Kindt M. Memory reconsolidation interference as an emerging treatment for emotional disorders: Strengths, limitations, challenges, and opportunities. Annu. Rev. Clin. Psychol 13 (2017): 99-121.

30. Kindt M, Soeter M, Vervliet B. Beyond extinction: erasing human fear responses and preventing the return of fear. Nat. Neurosci 12 (2009): 256-258.

31. Steenen SA, Wijk AJ, Heijden GJ, et al., Propranolol for the treatment of anxiety disorders: Systematic review and meta-analysis. J. Psychopharmacol (Oxf.) 30 (2015): 128-139.

32. Taylor JR, Torregrossa MM. Pharmacological Disruption of Maladaptive Memory, in: Kantak, K.M., Wettstein, J.G. (Eds.), Cognitive Enhancement, Handbook of Experimental Pharmacology. Springer International Publishing, Cham (2015): 381-415.

33. Soeter M, Kindt, M. Disrupting reconsolidation: pharmacological and behavioral manipulations. Learn. Mem 18 (2011): 357366.

34. Agren T, Engman J, Frick A, et al. Disruption of reconsolidation erases a fear memory trace in the human amygdala. Science 337 (2012): 1550-1552.

35. Björkstrand J, Agren T, Frick A, et al. Disruption of memory reconsolidation erases a fear memory trace in the human amygdala: An 18-month follow-up. PLoS ONE 10 (2015): 0129393.

36. Oyarzún JP, Lopez-Barroso D, Fuentemilla L, et al. Updating fearful memories with extinction training during reconsolidation: A human study using auditory aversive stimuli. PLoS ONE 7 (2012): e38849.

37. Chalkia A, Van Oudenhove L, Beckers T. Preventing the return of fear in humans using reconsolidation update mechanisms: A verification report of Schiller et al. (2010). Cortex 129 (2020b): 510-525.

38. Schiller D, Kanen JW, LeDoux JE, et al. Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. Proc. Natl. Acad. Sci 110 (2013): 20040–20045.

39. Soeter M, Kindt M. Dissociating response systems: Erasing fear from memory. Neurobiol. Learn. Mem 94 (2010): 30-41.

40. Haubrich, J., Bernabo, M., Baker, A.G., Nader, K. (2020). Impairments to consolidation, reconsolidation, and long-term memory maintenance lead to memory erasure. Annu. Rev. Neurosci 43 (2020): 297-314.

41. Clem RL, Huganir RL. Calcium-permeable AMPA receptor dynamics mediate fear memory erasure. Science 330 (2010): 1108–1112.

42. Jarome TJ, Kwapis JL, Werner CT, et al. The timing of multiple retrieval events can alter GluR1 phosphorylation and the requirement for protein synthesis in fear memory reconsolidation. Learn. Mem 19 (2012a): 300-306.

43. Agren T, Furmark T, Eriksson E, et al. Human fear reconsolidation and allelic differences in serotonergic and dopaminergic genes. Transl. Psychiatry 2 (2012): e76–e76.

44. Soeter M, Kindt M. Erasing fear for an imagined threat event. Psychoneuroendocrinology 37 (2012): 1769-1779.

45. Han DH, Park P, Choi DI, et al., The essence of the engram: Cellular or synaptic? Semin. Cell Dev. Biol (2021): S1084952121001488.

46. Ecker, B. Memory reconsolidation understood and misunderstood. Int. J. Neuropsychother 3 (2015a): 2-46.

47. Ecker, B. Clinical translation of memory reconsolidation research: Therapeutic methodology for transformational change by erasing implicit emotional learnings driving symptom production. Int. J. Neuropsychother 6 (2018) 1-92.

48. Vaz A, Ecker B. Memory reconsolidation in psychotherapy for severe perfectionism within borderline personality. J. Clin. Psychol 76 (2020): 2067-2078.

49. Elsey JWB, Van AstA, Kindt M. Human memory reconsolidation: A guiding framework and critical review of the evidence. Psychol. Bull 144 (2018): 797-848.

50. Zuccolo PF, Hunziker MHL. Preliminary study of the effects of post-retrieval extinction on the return of conditioned responses in humans. Rev. Bras. Análise Comportamento 14 (2018): 126-135.

51. Chalkia A, Weermeijer J, Van Oudenhove L, et al. Acute but not permanent effects of propranolol on fear memory expression in humans. Front. Hum. Neurosci 13 (2019): 51.

52. Chalkia, A, Schroyens N, Leng L, et al. No persistent attenuation of fear memories in humans: A registered replication of the reactivation-extinction effect. Cortex 129 (2020a): 496-509.

53. Kredlow MA, Orr SP, Otto MW. Exploring the boundaries of post-retrieval extinction in healthy and anxious individuals. Behav. Res. Ther 108 (2018): 45-57.

54. Schroyens N, Alfei JM, Schnell AE, et al. Limited replicability of drug-induced amnesia after contextual fear memory retrieval in rats. Neurobiol. Learn. Mem 166 (2019): 107105.

55. Auber A, Tedesco V, Jones CE, et al. Post-retrieval extinction as reconsolidation interference: methodological issues or boundary conditions? Psychopharmacology (Berl.) 226 (2013): 631-647.

56. Elsey JWB, Kindt M. Tackling maladaptive memories through reconsolidation: From neural to clinical science. Neurobiol. Learn. Mem 142 (2017): 108-117.

57. Nader K, Hard, O. A single standard for memory: the case for reconsolidation. Nat. Rev. Neurosci 10 (2009): 224-234.

58. Zhang JJ, Haubrich J, Bernabo M, et al. Limits on lability: Boundaries of reconsolidation and the relationship to metaplasticity. Neurobiol. Learn. Mem 154 (2018): 78-86.

59. Marks EH, Zoellner LA. Attenuating fearful memories: Effect of cued extinction on intrusions. Emotion 14 (2014): 1143-1154.

60. Björkstrand J, Agren T, Frick A, et al. Disrupting reconsolidation attenuates long-term fear memory in the

human amygdala and facilitates approach behavior. Curr. Biol 26 (2016): 2690-2695.

61. Björkstrand J, Agren T, Frick A, et al. Think twice, it's all right: Long lasting effects of disrupted reconsolidation on brain and behavior in human long-term fear. Behav. Brain Res 324 (2017): 125-129.

62. Feng P, Zheng Y, Feng T. Resting-state functional connectivity between amygdala and the ventromedial prefrontal cortex following fear reminder predicts fear extinction. Soc. Cogn. Affect. Neurosci 11 (2016): 991-1001.

63. Meir Drexler S, Merz CJ, Lissek S, et al. Reactivation of the unconditioned stimulus inhibits the return of fear independent of cortisol. Front. Behav. Neurosci 13 (2019): 254.

64. Yang Y, Jie J, Li J, et al. A novel method to trigger the reconsolidation of fear memory. Behav. Res. Ther 122 (2019): 103461.

65. Reichelt Amy C, Lee JLC. Over-expectation generated in a complex appetitive goal-tracking task is capable of inducing memory reconsolidation. Psychopharmacology (Berl.) 226 (2013): 649-658.

66. Bellfy L, Kwapis JL. Molecular mechanisms of reconsolidation-dependent memory updating. Int. J. Mol. Sci 21 (2020): 6580.

67. Campos-Arteaga G, Forcato C, Wainstein G, et al. Differential neurophysiological correlates of retrieval of consolidated and reconsolidated memories in humans: An ERP and pupillometry study. Neurobiol. Learn. Mem 174 (2020): 107279.

68. Ecker B, Ticic R, Hulley L. Unlocking the emotional brain: Memory reconsolidation and the psychotherapy of transformational change. Routledge, New York, NY (2024).

69. Dunbar AB, Taylor JR. (2017). Reconsolidation and psychopathology: Moving towards reconsolidation-based treatments. Neurobiol. Learn. Mem 142 (2017): 162-171.

70. Krawczyk MC, Fernández RS, Pedreira ME, et al. Toward a better understanding on the role of prediction error on memory processes: From bench to clinic. Neurobiol. Learn. Mem. 142 (2017): 13-20.

71. Kroes MCW, Schiller D, LeDoux JE, et al. Translational approaches targeting reconsolidation, in: Robbins, T.W., Sahakian, B.J. (Eds.), Translational Neuropsychopharmacology, Current Topics in Behavioral Neurosciences. Springer International Publishing, Cham (2015): 197-230.

72. Nader K, Hardt O, Lanius R. Memory as a new therapeutic target. Dialogues Clin. Neurosci. 15 (2014): 475-486.

73. Treanor M, Brown LA, Rissman J, et al. Can memories of traumatic experiences or addiction be erased or modified? A critical review of research on the disruption of memory reconsolidation and its applications. Perspect. Psychol. Sci 12 (2017): 290-305.

74. Thompson A, Lipp OV. Extinction during reconsolidation eliminates recovery of fear conditioned to fear-irrelevant and fear-relevant stimuli. Behav. Res. Ther. 92 (2017): 1-10.

75. Baker KD, McNally GP, Richardson R. Memory retrieval before or after extinction reduces recovery of fear in adolescent rats. Learn. Mem 20 (2013): 467-473.

76. Millan EZ, Milligan-Saville J, McNally GP. Memory retrieval, extinction, and reinstatement of alcohol seeking. Neurobiol. Learn. Mem 101 (2013): 26-32.

77. Ponnusamy R, Zhuravka I, Poulos AM, Shobe J, et al. Fanselow, M.S. Retrieval and reconsolidation accounts of fear extinction. Front. Behav. Neurosci 10 (2016).

78. Lee JLC. Reconsolidation: maintaining memory relevance. Trends Neurosci 32 (2009): 413-420.

79. Díaz-Mataix L, Debiec J, LeDoux JE, Doyère V. Sensory specific associations stored in the lateral amygdala allow for selective alteration of fear memories. J. Neurosci 31 (2011): 9538-9543.

80. Kim J, Song B, Hong I, et al. Reactivation of fear memory renders consolidated amygdala synapses labile. J. Neurosci 30 (2010): 9631-9640.

81. Maren S. Seeking a Spotless Mind: Extinction, deconsolidation, and erasure of fear memory. Neuron 70 (2011): 830-845.

82. Pedreira ME, Pérez-Cuesta LM, Maldonado H. Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. Learn. Mem 11 (2004): 579-585.

83. Forcato C, Argibay PF, Pedreira ME, et al. Human reconsolidation does not always occur when a memory is retrieved: The relevance of the reminder structure. Neurobiol. Learn. Mem 91 (2009): 50-57.

84. Sevenster D, Beckers T, Kindt M. Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. Neurobiol. Learn. Mem 97 (2012): 338-345.

85. Grégoire L, Greening SG. Opening the reconsolidation window using the mind's eye: Extinction training during reconsolidation disrupts fear memory expression following mental imagery reactivation. Cognition 183 (2019): 277-281.

86. Cahill EN, Milton AL. Neurochemical and molecular

mechanisms underlying the retrieval-extinction effect. Psychopharmacology (Berl.) 236 (2019): 111-132.

87. Exton-McGuinness MTJ, Lee JLC, Reichelt AC. Updating memories: The role of prediction errors in memory reconsolidation. Behav. Brain Res. 278 (2015): 375-384.

88. Fernández RS, Boccia MM, Pedreira ME. The fate of memory: Reconsolidation and the case of Prediction Error. Neurosci. Biobehav. Rev 68 (2016b): 423-441.

89. Rodriguez-Ortiz CJ, Bermúdez-Rattoni F. Determinants to trigger memory reconsolidation: The role of retrieval and updating information. Neurobiol. Learn. Mem 142 (2017): 4-12.

90. Sinclair AH, Barense MD. Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. Trends Neurosci 42 (2019): 727-739.

91. Agren T. Human reconsolidation: A reactivation and update. Brain Res. Bull 105 (2014): 7082.

92. Delorenzi A, Maza FJ, Suárez LD, et al. Memory beyond expression. J. Physiol.-Paris 108 (2014): 307–322.

93. Boccia MM. Post-retrieval effects of icv infusions of hemicholinium in mice are dependent on the age of the original memory. Learn. Mem 13 (2006): 376-381.

94. Debiec J, LeDoux JE, Nader K. Cellular and systems reconsolidation in the hippocampus. Neuron 36 (2002): 527-538.

95. Debiec J, Diaz-Mataix L, Bush DEA, et al. The selectivity of aversive memory reconsolidation and extinction processes depends on the initial encoding of the Pavlovian association. Learn. Mem 20 (2013): 695-699.

96. Eisenberg M, Dudai Y. Reconsolidation of fresh, remote, and extinguished fear memory in medaka: old fears don't die. Eur. J. Neurosci 20 (2004): 3397-3403.

97. Frankland PW. Stability of recent and remote contextual fear memory. Learn. Mem 13 (2006): 451-457.

98. Inda MC, Muravieva EV, Alberini, CM. Memory retrieval and the passage of time: From reconsolidation and strengthening to extinction. J. Neurosci 31 (2011): 1635-1643.

99. Milekic MH, Alberini CM. Temporally graded requirement for protein synthesis following memory reactivation. Neuron 36 (2002): 521-525.

100. Steinfurth ECK, Kanen JW, Raio CM, et al. (2014). Young and old Pavlovian fear memories can be modified with extinction training during reconsolidation in humans. Learn. Mem. 21 (2014): 338-341.

101. Suzuki A, Josselyn SA, Frankland PW, et al. Memory reconsolidation and extinction have distinct temporal and biochemical signatures. J. Neurosci 24 (2004): 4787-4795.

102. Winters BD, Tucci MC, DaCosta-Furtado M. Older and stronger object memories are selectively destabilized by reactivation in the presence of new information. Learn. Mem 16 (2009): 545-553.

103. Kida S. Function and mechanisms of memory destabilization and reconsolidation after retrieval. Proc. Jpn. Acad. Ser. B 96 (2020): 95-106.

104. Schwabe L, Nader K, Pruessner JC. Reconsolidation of human memory: Brain mechanisms and clinical relevance. Biol. Psychiatry 76 (2014): 274-280.

105. Monfils MH, Holmes EA. Memory boundaries: opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. Lancet Psychiatry 5 (2018): 1032-1042.

106. Hernandez PJ, Kelley AE. Long-term memory for instrumental responses does not undergo protein synthesis-dependent reconsolidation upon retrieval. Learn. Mem 11 (2004): 748-754.

107. Wang SH. Consolidation and Reconsolidation of Incentive Learning in the Amygdala. J. Neurosci 25 (2005): 830-835.

108. Bos MGN, Beckers T, Kindt M. Noradrenergic blockade of memory reconsolidation: A failure to reduce conditioned fear responding. Front. Behav. Neurosci 8 (2014): 412.

109. Alfei JM, Ferrer Monti RI, Molina VA, et al. Prediction error and trace dominance determine the fate of fear memories after post-training manipulations. Learn. Mem 22 (2015): 385-400.

110. Chen W, Li J, Xu L, et al. Destabilizing different strengths of fear memories requires different degrees of prediction error during retrieval. Front. Behav. Neurosci 14 (2021a): 598924.

111. Díaz-Mataix L, Ruiz Martinez RC, Schafe G E, et al. Detection of a temporal error triggers reconsolidation of amygdala-dependent memories. Curr. Biol 23 (2013): 467-472.

112. Jarome TJ, Kwapis JL, Werner CT, et al. The timing of multiple retrieval events can alter GluR1 phosphorylation and the requirement for protein synthesis in fear memory reconsolidation. Learn. Mem 19 (2012b): 300-306.

113. Li J, Chen W, Caoyang J, et al. Moderate partially reduplicated conditioned stimuli as retrieval cue can increase effect on preventing relapse of fear to compound stimuli. Front. Hum. Neurosci 11 (2017): 575.

114. Piñeyro ME, Ferrer Monti RI, Alfei JM, et al. Memory destabilization is critical for the success of the reactivation-extinction procedure. Learn. Mem 21 (2014): 46-54.

115. Sevenster D, Beckers T, Kindt M. Prediction error governs pharmacologically induced amnesia for learned fear. Science 339 (2013): 830-833.

116. Sevenster D, Beckers T, Kindt M. Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. Learn. Mem 21 (2014): 580-584.

117. Junjiao L, Wei C, Jingwen C, et al. Role of prediction error in destabilizing fear memories in retrieval extinction and its neural mechanisms. Cortex 121 (2019): 292-307.

118. Cahill EN, Wood MA, Everitt BJ, et al. The role of prediction error and memory destabilization in extinction of cued-fear within the reconsolidation window. Neuropsychopharmacology 44 (2019): 1762-1768.

119. Cochran A L, Cisler J M. A flexible and generalizable model of online latent-state learning. PLoS Comput Biol 15 (2019): e1007331.

120. Courville A C, Daw N D, Touretzky D S. Bayesian theories of conditioning in a changing world. Trends Cogn. Sci 10 (2006): 294-300.

121. Gershman SJ, Monfils MH, Norman KA, et al. The computational nature of memory modification. eLife 6 (2017): e23763.

122. Gershman SJ, Jones CE, Norman KA, et al. Gradual extinction prevents the return of fear: implications for the discovery of state. Front. Behav. Neurosci 7 (2013): 164.

123. Kennedy NGW, Lee JC, Killcross S, et al. Prediction error determines how memories are organized in the brain. eLife 13 (2024): RP95849.

124. López MA, Santos MJ, Cortasa S, et al. Different dimensions of the prediction error as a decisive factor for the triggering of the reconsolidation process. Neurobiol. Learn. Mem 136 (2016): 210-219.

125. Asthana MK, Brunhuber B, Mühlberger A, et al. Preventing the return of fear using reconsolidation update mechanisms depends on the met-allele of the brain derived neurotrophic factor Val66Met polymorphism. Int. J. Neuropsychopharmacol 19 (2016): 1-9.

126. Pine A, Mendelsohn A, Dudai Y. Unconscious learning of likes and dislikes is persistent, resilient, and reconsolidates. Front. Psychol 5 (2014): 1051.

127. Chen W, Li J, Zhang X, et al. Retrieval-extinction as a reconsolidation-based treatment for emotional disorders:Evidence from an extinction retention test shortly after intervention. Behav. Res. Ther 139 (2021b): 103831.

128. Kitamura H, Johnston P, Johnson L, et al. Boundary conditions of post-retrieval extinction: A direct comparison of low and high partial reinforcement. Neurobiol. Learn. Mem 174 (2020): 107285.

129. Houtekamer M C, Henckens MJAG, Mackey WE, et al. Investigating the efficacy of the reminder-extinction procedure to disrupt contextual threat memories in humans using immersive Virtual Reality. Sci. Rep 10 (2020): 16991.

130. Ecker B. Using NLP for memory reconsolidation: A glimpse of integrating the panoply of psychotherapies. The Neuropsychotherapist 10 (2015b): 50-56.

131. Golkar A, Bellander M, Olsson A, et al. Are fear memories erasable? Reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. Front. Behav. Neurosci 6 (2012): 80.

132. Kroes MCW, Dunsmoor JE, Lin Q, et al. A reminder before extinction strengthens episodic memory via reconsolidation but fails to disrupt generalized threat responses. Sci. Rep 7 (2017): 10858.

133. Chen W, Lin X, Li J, et al. Gender difference in retrieval-extinction of conditioned fear memory. Acta Psychol. Sin 53 (2021c): 1082.

134. Liu J, Zhao L, Xue Y, et al. An unconditioned stimulus retrieval extinction procedure to prevent the return of fear memory. Biol. Psychiatry 76 (2014): 895-901.

135. Debiec J, Díaz-Mataix L, Bush DEA, et al. The amygdala encodes specific sensory features of an aversive reinforcer. Nat. Neurosci 13 (2010): 536-537.

136. Johnson DC, Casey BJ. Extinction during memory reconsolidation blocks recovery of fear in adolescents. Sci. Rep 5 (2015): 1-5.

137. Phelps EA, Hofmann SG. Memory editing from science fiction to clinical practice. Nature 572 (2019): 43-50.

138. Haubrich J, Crestani AP, Cassini LF, et al. Reconsolidation allows fear memory to be updated to a less aversive level through the incorporation of appetitive information. Neuropsychopharmacology 40 (2015): 315-326.

139. Olshavsky ME, Song BJ, Powell DJ, et al. Updating appetitive memory during reconsolidation window: critical role of cue-directed behavior and amygdala central nucleus. Front. Behav. Neurosci. 7 (2013).

140. Redondo RL, Kim J, Arons AL, et al. Bidirectional switch of the valence associated with a hippocampal

contextual memory engram. Nature 513 (2014): 426-430.

141. Hutton-Bedbrook K, McNally GP. The promises and pitfalls of retrieval-extinction procedures in preventing relapse to drug seeking. Front. Psychiatry 4 (2013): 14.

142. Cai DJ, Aharoni D, Shuman T, et al. A shared neural ensemble links distinct contextual memories encoded close in time. Nature 534 (2016): 115-118.

143. Rashid AJ, Yan C, Mercaldo V, et al. Competition between engrams influences fear memory formation and recall. Science 353 (2016): 383-387.

144. Phelps EA, Delgado MR, Nearing KI, et al. Extinction learning in humans: Role of the amygdala and vmPFC. Neuron 43 (2004): 897-905.

145. Santini E, Ge H, Ren K, et al. Consolidation of fear extinction requires protein synthesis in the medial prefrontal cortex. J. Neurosci 24 (2004): 5704-5710.

146. Forcato C, Rodríguez MLC, Pedreira ME, et al. Reconsolidation in humans opens up declarative memory to the entrance of new information. Neurobiol. Learn. Mem. 93 (2010): 77-84.

147. Hupbach A, Gomez R, Hardt O, et al. Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. Learn. Mem 14 (2007): 47-53.

148. Warren VT, Anderson KM, Kwon C, et al. Human fear extinction and return of fear using reconsolidation update mechanisms: The contribution of on-line expectancy ratings. Neurobiol. Learn. Mem 113 (2014): 165-173.

149. Meir Drexler S, Merz CJ, Hamacher-Dang TC, et al. Effects of postretrieval-extinction learning on return of contextually controlled cued fear. Behav. Neurosci 128 (2014): 474-481.

150. Hu J, Wang W, Homan P, et al. Reminder duration determines threat memory modification in humans. Sci. Rep 8 (2018): 8848.

151. Harvey AG, Lee J, Williams J, et al. Improving outcome of psychosocial treatments by enhancing memory and learning. Perspect. Psychol. Sci 9 (2014): 161-179.

152. Shedler J. Where is the evidence for "evidence-based" therapy? J. Psychol. Ther. Prim. Care 4 (2015): 47-59.

153. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. Am. Psychol 32 (1977): 752-760.

154. Wampold BE, Imel ZE. The great psychotherapy debate: The evidence for what makes psychotherapy work. Routledge, New York (2015).

155. Ecker B, Ticic R, Hulley L. Unlocking the Emotional Brain: Eliminating Symptoms at Their Roots Using Memory Reconsolidation. Routledge, New York, NY (2012).

156. Högberg G. Affective psychotherapy in post-traumatic reactions guided by affective neuroscience: memory reconsolidation and play. Psychol. Res. Behav. Manag 87 (2011): 87-96.

157. Cuijpers P, Reijnders M, Huibers MJH. The role of common factors in psychotherapy outcomes. Annu. Rev. Clin. Psychol 15 (2019): 207-231.

158. Goldfried MR. Obtaining consensus in psychotherapy: What holds us back? Am. Psychol 74 (2019): 484-496.

159. Kazdin AE. Understanding how and why psychotherapy leads to change. Psychother. Res 19 (2009): 418-428.

160. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. Annu. Rev. Clin. Psychol 3 (2007): 1-27.

161. Barreiro KA, Suárez LD, Lynch VM, et al. Memory expression is independent of memory labilization/reconsolidation. Neurobiol. Learn. Mem 106 (2013): 283-291.

162. Santoyo-Zedillo M, Rodriguez-Ortiz CJ, Chavez-Marchetta G, et al. Retrieval is not necessary to trigger reconsolidation of object recognition memory in the perirhinal cortex. Learn. Mem 21 (2014): 452-456.

163. Moyal N, Henik A, Anholt G. Cognitive strategies to regulate emotions—Current evidence and future directions. Frontiers in Psychology 4 (2014): 1019.

164. Ochsner K N, Gross JJ. The cognitive control of emotion. Trends Cogn. Sci 9 (2005): 408-409.

165. Toomey B, Ecker B. Competing visions of the implications of neuroscience for psychotherapy. J. Constr. Psychol 22 (2009): 95-140.

166. Ecker B, Hulley L. Coherence Therapy Practice Manual and Training Guide. Albany, California: Coherence Psychology Institute (2019).

167. Herink R. The psychotherapy handbook. New American Library, New York (1980).

168. Lane RD, Ryan L, Nadel L, et al. Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. Behav. Brain Sci 38 (2015): e1

169. Ecker B, Hulley L, Ticic R. Minding the findings: Let's not miss the message of memory reconsolidation research for psychotherapy. Behav. Brain Sci 38 (2015): e7

170. Ecker B. Erasing problematic emotional learnings:

Psychotherapeutic use of memory reconsolidation research, in: Lane, R.D., Nadel, L. (Eds.), Neuroscience of Enduring Change: Implications for Psychotherapy. Oxford University Press (2020): 273-299.

171. Stevens FL. Affect regulation and affect reconsolidation as organizing principles in psychotherapy. J. Psychother. Integr 29 (2019): 277-290.

172. Welling H. Transformative emotional sequence: Towards a common principle of change. J. Psychother. Integr 22 (2012): 109-136.

173. Alberini CM. Commentary on Tuch. J. Am. Psychoanal. Assoc 63 (2015): 317-330.

174. Walker Matthew P, Brakefield T, Allan Hobson J, et al. Dissociable stages of human memory consolidation and reconsolidation. Nature 425 (2003): 616-620.

175. Goldman RN, Greenberg LS, Pos AE. Depth of emotional experience and outcome. Psychother. Res 15 (2005): 248-260.

176. Greenberg L S. Emotion-focused therapy: A clinical synthesis. Focus 8 (2010): 32-42.

177. Dominguez E, Casagrande M, Raffone A. Autobiographical memory and mindfulness: A critical review with a systematic search. Mindfulness (2022): 1614-1651.

178. Kerswell NL, Strodl E, Hawkins D, et al. Memory reconsolidation therapy for police officers with post-traumatic stress disorder. J. Police Crim. Psychol 36 (2021): 112-123.

179. Kindt M, Soeter M. Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. Biol. Psychol 92 (2013): 43-50.

180. Chen W, Li J, Caoyang J, et al. Effects of prediction error on post-retrieval extinction of fear to compound stimuli. Acta Psychol. Sin 50 (2018): 739-749.

181. Coppens E, Spruyt A, Vandenbulcke M, et al. Classically conditioned fear responses are preserved following unilateral temporal lobectomy in humans when concurrent US-expectancy ratings are used. Neuropsychologia 47 (2009): 2496-2503.

182. Drysdale AT, Hartley CA, Pattwell SS, et al. Fear and anxiety from principle to practice: Implications for when to treat youth with anxiety disorders. Biol. Psychiatry 75 (2014): 19-20.

183. Eisenberg M, Kobilo T, Berman DE, et al. Stability of retrieved memory: Inverse correlation with trace dominance. Science 301 (2003): 1102-1104.

184. Fernandez-Rey J, Gonzalez-Gonzalez D, Redondo J. Preventing the return of fear memories with postretrieval extinction: A human study using a burst of white noise as an aversive stimulus. Behav. Neurosci 132 (2018): 230-239.

185. Forcato C, Fernandez R S, Pedreira M E. Strengthening a consolidated memory: The key role of the reconsolidation process. J. Physiol.-Paris 108 (2014): 323-333.

186. Fricchione J, Greenberg MS, Spring J, et al. Delayed extinction fails to reduce skin conductance reactivity to fear-conditioned stimuli: Delayed extinction fails to reduce reactivity. Psychophysiology 53 (2016): 1343-1351.

187. Funayama ES, Grillon C, Davis M, et al. A double dissociation in the affective modulation of startle in humans: Effects of unilateral temporal lobectomy. J. Cogn. Neurosci 13 (2001): 721-729.

188. Golkar A, Tjaden C, Kindt M. Vicarious extinction learning during reconsolidation neutralizes fear memory. Behav. Res. Ther 92 (2017): 87-93.

189. Greenberg LS. Emotions, the great captains of our lives: Their role in the process of change in psychotherapy. Am. Psychol 67 (2012): 697-707.

190. Klucken T, Kruse O, Schweckendiek J, et al. No evidence for blocking the return of fear by disrupting reconsolidation prior to extinction learning. Cortex 79 (2016): 112-122.

191. Lee JLC, Milton AL, Everitt BJ. Reconsolidation and extinction of conditioned fear: Inhibition and potentiation. J. Neurosci 26 (2006): 10051-10056.

192. Luo Q, Holroyd T, Majestic C, et al. Emotional automaticity is a matter of timing. J. Neurosci 30 (2010): 5825-5829.

193. Meir Drexler S, Wolf OT. Behavioral disruption of memory reconsolidation: From bench to bedside and back again. Behav. Neurosci 132 (2018b): 13-22.

194. Pattwell SS, Duhoux S, Hartley CA, et al. Altered fear learning across development in both mouse and human. Proc. Natl. Acad. Sci 109 (2012): 16318-16323.

195. Rossato JI, Bevilaqua LRM, Myskiw JC, et al. On the role of hippocampal protein synthesis in the consolidation and reconsolidation of object recognition memory. Learn. Mem 14 (2007): 36-46.

196. Schiller D, LeDoux JE, Phelps EA. Reply to Beckers, McIntosh and Chambers on the verification of 'preventing the return of fear using retrieval-extinction in humans' (preprint) PsyArXiv (2020).

197. Zimmermann J, Bach DR. Impact of a reminder/ extinction procedure on threat-conditioned pupil size and skin conductance responses. Learn. Mem 27 (2020): 164-172.

Appendix A. 20 successful studies of human behavioral updating

The following reviews provide details of each study's procedure and results that are relevant to identifying the experiences induced in subjects, including prediction error (PE) experiences. For full details, the original journal article must be consulted. Examples of items not included here are baseline measurement procedures, habituation procedures, statistical analyses, subject demographics, and subject inclusion and exclusion criteria. Acronyms are as defined with Table 1.

The PE analyses of these studies range from somewhat speculative to strongly probable. For any one PE analysis, some other analysis certainly could be argued. However, the purpose of each PE analysis is not to argue for its final veracity, but only to explore the possibility that the well-defined set of PE analysis criteria delineated in Section 3 could logically and consistently account for all results of experiments on behavioral updating of human emotional memory.

A.1. Schiller et al. (2010) [6]: "Preventing the return of fear in humans using reconsolidation update mechanisms"

The main elements of the two retrieval-extinction protocols used by Schiller et al. [6] are described in Section 3.3. Fear responses were measured as skin conductance response (SCR). In the first of the two reported experiments, Day 1 had fear acquisition training consisting of Pavlovian classical conditioning, with the CSs consisting of squares of two colors (yellow and blue) presented in random order on a computer screen. One color served as the conditioned stimulus CS+ paired with an electric shock US, with a 38% reinforcement schedule (6 of 16 CS+ presentations). The other colored square, CS–, presented 10 times, was never accompanied by a shock. Each CS square was onscreen for 4 s, with an inter-trial interval (ITI) of 11±1 s.

On Day 2, subjects experienced cued reactivation of the CS+ fear response by a single unreinforced presentation of CS+ for 4 s, then saw a "television show episode" appear on the computer screen for 10 min, and then experienced a random series of unreinforced presentations of 10 CS+ and 11 CS– with the same ITI as on Day 1. That series is referred to as "extinction" by Schiller et al., and indeed it would be a standard extinction procedure if it had occurred by itself, but, as explained below, by coming after the prior steps of the Day 2 sequence, its effects differed fundamentally from those of standard extinction, so calling it "extinction" could cause misconceptions.

On Day 3, spontaneous recovery was measured via a re-extinction process consisting again of unreinforced presentations of 10 CS+ and 11 CS–. Annulment of the acquired fear memory for CS+ was indicated by measuring no statistical difference between SCR levels for CS+ and CS– in the first trial of the re-extinction procedure. Annulment persisted in a follow-up after one year, when SCR levels in response to CS+ and CS– were again equal in a test of reinstatement, consisting of four unsignalled US shocks followed by presentations of unreinforced CS+ and CS–.

In contrast, fear of CS+ returned on Day 3 for two other groups of subjects, one with the 10-min break removed, making the procedure a standard extinction training, and the other with a break of 6 h instead of 10 min between retrieval and extinction on Day 2, so that the CS+ fear memory was no longer destabilized when extinction was implemented. In Day 3 spontaneous recovery testing, SCR levels for CS+ were significantly higher than for CS– in those two groups.

By 2010, the requirement of a PE experience for inducing memory destabilization and reconsolidation had been reported by 12 studies published from 2004 to 2009 (see https://bit.ly/2b8IbJH for a chronological list of over 30 studies that have confirmed the PE requirement) and had been a primary focus of a review article [78]. However, the requirement of a PE experience for inducing memory destabilization is not considered in the interpretation of results given by Schiller et al. [6]. Rather, they infer that "timing may have a more important role in the control of fear than previously appreciated. …Our findings indicate that the timing of extinction relative to the reactivation of the memory can capitalize on reconsolidation mechanisms" (p. 52). That seems to suggest that the 10-min time delay was somehow intrinsic for updating a fear memory through reconsolidation.

Subsequently it became widely assumed among researchers that there exists an intrinsic requirement for a time delay in order to induce target memory destabilization. For example, the authors of one study [178] wrote that their clinical protocol "incorporates memory recall followed by a wait period to make memories labile for modification" (p. 112). Other researchers anticipating clinical application [136] likewise reasoned (p. 4), "A modified version of an exposure-based CBT protocol based on memory reconsolidation might involve reminding patients of why they are there when they first arrive at the clinician's office (i.e., reminder cue), then establishing a safe and positive rapport for approximately 10 minutes (i.e., waiting for reconsolidation window) before initiating desensitization with exposure therapy."

Viewing the 10-min delay as an intrinsic requirement for MR updating of fear memory was only a speculation made by Schiller et al. [6] without consideration of the PE requirement, but was widely taken as firmly established. A quite different interpretation of this study's results has been suggested based on analyzing how the procedure of Schiller et al. created a PE experience [46], as is done again in Section 3.3.1 above. The latter analysis distills down to the question of how the Day 2 transition from the offset (disappearance) of a single reminder CS+ presentation to the unexpected appearance onscreen of a "television show episode" was made.

If the transition occurred in such a way that subjects regarded the viewing of the TV episode as a true part of the study, a PE experience would be created, but if it was regarded as occurring outside of the study, no PE experience would occur. Several different elements in the transition could tip that construal one way of the other. For example, leaving all electronics attached to the subject's body and turned on would tend to be perceived as meaning that the TV presentation is part of the study, unless other communications more strongly indicated the opposite. Disconnection of the electronics (as was done by, for example, in the studies reviewed in Sections B.3, B.7 and B.11, each of which reported replication failure) would tend to mean the 10-min period is not part of the study. Researchers' verbal communication to subjects about the transition and the 10-min period could likewise be strongly influential one way of the other. In their descriptions of procedure, Schiller et al. [6] repeatedly refer to the 10-min period as a "10-min break," but whether that phrase, which would tend to imply that the TV episode viewing is outside of the study, was ever used in communicating with subjects is not divulged, and neither is any other aspect of how the transition was actually carried out. It is possible that the transition could comprise a totality of signifiers with enough complexity and inconsistency that, within a single study, some subjects would view the 10-min period as being inside the study, some would view it as being outside the study, and some would be confused about whether it was inside or outside of the study.

Therefore, PE analysis of this study is in a relatively weak position and must consist of applying the PE inference principle defined in Section 3.2.1: The fact that updating and annulment of threat memory resulted for the reminder group means that a PE experience was created, and the only possible point in the procedure that could have produced that PE experience was the unexpected appearance of the TV show episode, therefore it must be that subjects were regarding the TV show episode and the 10-min period as occurring within the study, as a part of the study. Otherwise, no PE experience would have occurred, and no updating.

That is a reverse-engineering analysis, and it is reminiscent of the famous words that author Arthur Conan Doyle put in the mouth of Sherlock Holmes: "Once you eliminate the impossible, whatever remains, no matter how improbable, must be the truth." This is not circular reasoning, rather it is an exploration of (a) whether it is conceptually possible to account for all retrieval-extinction experimental results in terms of the presence or absence of PE experiences, and (b) how PE phenomenology has to be modeled to be a comprehensive fit. That modeling then requires empirical confirmation by testing of its predictions. The analysis of PE experience in numerous other studies in Appendices A and B is on a significantly firmer basis.

Another factor that probably favored creation of a PE experience by the TV episode in this experiment is the uncertainty of the target memory latent cause. This factor is discussed in Section 3.3.1. A strongly uncertain latent cause is indicated in this study by (a) the 38% acquisition reinforcement, which produces strong uncertainty of US occurrence, and (b) the single unreinforced CS+ reactivation immediately preceding the TV show, which was ambiguous as to whether its latent cause was the same or different from the latent cause of the acquisition training. Thus the TV show appeared at a moment when latent cause uncertainty was strong, so the TV show would be perceived as not definitely irrelevant to that latent cause, i.e., possibly relevant to it, preserving that uncertain latent cause and generating a PE experience, destabilizing the target memory.

The TV show's duration of 10 min was presumably more of the same novelty as seeing the TV show appear, not a separate novelty generating yet another PE experience. If TV show duration had been 11 s (same duration as the expected blank screen) instead of 10 min, that would have produced the same PE experience, causing destabilization and allowing subsequent fear memory annulment, by this analysis. That testable prediction challenges the interpretation by Schiller et al. [6] that the 10-min time delay after reactivation is an inherent requirement for fear memory updating. The inherent requirement is a PE experience. Once a PE experience has occurred, the neurophysiological and neurochemical process of destabilization presumably requires some minutes to initiate sufficiently, but probably much less than 10 min.

It is instructive to analyze also the hypothetical case where subjects' activity during the 10-min interval on Day 2 would consist of simply resting and waiting with the computer screen blank. Waiting with nothing happening on the screen is what all subjects learned on Day 1 to expect after seeing a CS image, so after offset of the first CS+ presentation on Day 2, PE would not be created by seeing a blank screen at first. However, well before 10 min, the duration of the blank screen would sufficiently violate the expectation of an 11-s inter-trial interval to produce PE and destabilization. That prediction is reliable because a purely temporal mismatch of the acquisition training's timing structure is known to induce memory destabilization [109, 111].

Thus, in the retrieval-extinction protocol, depending on what is or is not onscreen during the Day 2 10-min interval, different types of PE can be generated, and at different time points. In any case, according to this framework of PE analysis, a post-reactivation time delay per se would generate PE and destabilization merely because the target memory contains well-defined timing expectations that are violated by the delay. The 10-min delay used by Schiller et al. [6] is therefore not an inherent aspect of fear updating by the MR mechanism.

These considerations illustrate the utility of the mismatch and PE analysis framework in Section 3 for identifying cause and effect in procedures intended to induce destabilization and reconsolidation. With the target learning destabilized by the unexpected TV show following the first CS+ presentation on Day 2, the ensuing series of 10 unreinforced CS+ presentations functioned as counter-learning that directly updated the target memory's expectation that shock will accompany 38% of CS+ squares to an expectation that no shock will accompany any CS+ square. Fear disappeared as a result of that fundamental change wrought in the expectational mental model.

Most (but not all) of the successful human studies listed in Table 1 followed one of the two retrieval-extinction protocols used by Schiller et al. [6] with specialized variations. That attests to at least the qualitative validity of the results reported by Schiller et al., despite subsequent challenges to the validity of how their data was processed [37] (for the rebuttal, see [196]).

In summary, annulment of a learned fear was accomplished in experiment 1 by Schiller et al. [6] through this set of three experiences:

- experience of the reactivated target memory;
- presumably, a PE experience in relation to the reactivated target memory;
- soon thereafter, a counter-learning experience contradicting how the target memory expects the world to be.

Emphasis here is on viewing annulment of the target emotional learning as being caused by combining those three types of experiences, rather than by the details any particular procedures used for creating those experiences. The details of procedure will vary greatly across the remaining 19 successful behavioral annulment studies reviewed below, but those three categories of experience will remain invariant. Therefore, it is proposed that that set of three experiences, not any particular concrete procedure, is what would serve as the best methodology template for achieving emotional memory annulment by both researchers and clinicians.

In experiment 2 described by Schiller et al. [6], a third colored square was added to the acquisition training, and two of the colors, CSa+ and CSb+, received the same 38% reinforcement by the electric shock US, although now on 5 of 13 presentations, plus 8 CS– presentations never paired with shock. On Day 2, the CSa+ and the CS–, but not the CSb+, were presented once, unreinforced, before the 10-min TV show appeared, followed by the extinction training with all three CSs. On Day 3, fear did not return upon seeing the CSa+, but did return upon seeing the CSb+, showing that the reactivation and destabilization processes permit a high degree of specificity and separability among even closely associated cues, which bodes well for the therapeutic use of the annulment process.

As noted, the series of 10 unreinforced CS presentations on Day 2 was referred to as "extinction" by Schiller et al. [6], resulting in the protocol universally being called retrieval-extinction or post-retrieval extinction. However, the actual functioning and effects of that series on Day 2, producing memory updating and annulment, differ fundamentally from standard extinction's behavioral effects, memory encoding effects, neural network engagement effects, and molecular cascade effects, as numerous studies have shown (reviewed in [46]); and again demonstrated recently by comparative fMRI imagining [11]. Many experimental results indicate that updating directly modifies the target memory [24], whereas extinction creates a separate memory that competes with the target memory [16]. One of the conclusions of a study by Duvarci and Nader [19] was, "Reconsolidation cannot be reduced down to facilitated extinction." It has therefore been argued [46] that "extinction" in the name of this protocol is a major misnomer and a potential source of misunderstanding. More appropriate names would be retrieval-updating, post-reactivation counter-learning, or post-reactivation unlearning, for example. Furthermore, there is no inherent necessity for the counter-learning experience to have the same procedural structure as conventional extinction (a series of many identical CS–noUS presentations). After destabilization has occurred, unlearning, updating, and annulment of the target learning should be achievable by any type of distinct experience that contradicts the target learning's expectations. Extinction is not used in one of the successful updating studies [126] reviewed below in Section A.7, illustrating this procedural freedom.

A.2. Oyarzún et al. (2012) [36]: "Updating fearful memories with extinction training during reconsolidation: A human study using auditory aversive stimuli"

Oyarzún et al. [36] contributed the first successful replication of the study by Schiller et al. [6]. It is patterned on experiment 2 reported by Schiller et al., using three different-color squares as CSs (yellow, pink, and blue), and it incorporates significant variations of procedure: During acquisition on Day 1, two of the CSs, CSa+ and CSb+, each again received 38% reinforcement, but by different USs: USa and USb, respectively, which were different aversive sounds (a girl screaming and a pig squealing), i.e., auditory USs rather than the electric shock used by Schiller et al.

The authors explain (p. 3) that "a different US for each specific CS allowed us to increase the CS-US specificity in order to prevent a single US from recalling the memory of both CSs during reactivation." The third color square was never paired with a US, so the authors refer to it as NS (neutral stimulus). Each square was presented for 4 s with an inter-trial interval of 9±1 s. The number of presentations, 10 for each CS, was the same as in Schiller et al. [6], as was the way of measuring fear, namely SCR.

The procedure on Day 2 replicated that of Schiller et al. [6], beginning for all subjects with memory reactivation via one unreinforced presentation of CSa+ and one presentation of NS, followed by the appearance of a TV show video on the screen for 10 min, followed by extinction (randomized, unreinforced presentations of CSa+, CSb+, and NS, 10 of each). With the same 38% reinforcement at Day 1 acquisition and the same Day 2 reactivation conditions as those of Schiller et al., the PE analysis is also the same. Oyarzún et al. [36] indicate recognition of the PE requirement by stating, "in order to induce reconsolidation, the reminder should generate a mismatch between what is expected and what actually happens" (p. 6).

On Day 3 the test for recovery versus elimination of fear consisted of reinstatement by four unsignalled US presentations, followed after 10 min by measuring SCR in response to re-extinction (unreinforced CS presentations). In that SCR data, fear of the Day 2 unreminded CSb+ returned with strength equal to that produced by acquisition on Day 1, but fear of the reminded CSa+ did not increase above its extinguished level at the end of Day 2, reproducing the nullification of threat memory reported by Schiller et al. [6].

Though the procedure of this study differed significantly from that of Schiller et al. [6] (by using two different USs in a different sensory mode), the procedure created the same set of three experiences: target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.3. Agren et al. (2012) [34]: "Disruption of reconsolidation erases a fear memory trace in the human amygdala"

As implemented by Agren et al. [34], the retrieval-extinction protocol underwent further variations. During acquisition on Day 1, each human subject viewed a computer screen and saw the image of a neutral scene containing a lamp. Every 20 s, the lamp-light turned on and was either red or blue for 6 s, a total of 32 times, 16 in each color in random order. Between lightings, the lamp was unlit in the same scene for 14 s. One of the lamp colors, CS+, was 100% reinforced by a US, an unpleasant electric shock to the lower right arm, delivered 0.25 sec before the lamp light turned off. In that way subjects acquired a learned fear response to one color but not the other, including the expectation that shock would always accompany the feared color. Skin conductance response (SCR) measurements detected the fear response physiologically.

That acquisition learning was the target for unlearning and annulment by the procedure carried out on Day 2. Agren et al. [34], like Schiller et al. [6], never refer to the requirement of PE for producing destabilization. The PE analysis here, again an application of the framework in Section 3, is relatively complex. Subjects viewed the same computer screen on Day 2 and first saw the CS+ image of the shock-associated lamp color appear for 2 min, with no shock occurring. The target learning was reactivated immediately by seeing the lamp light with that color. The lit lamp persisted for 2 min and then went dark with no shock occurring. The 2 min duration, being far longer than the 6-s presentation expected by the acquisition training memory, would launch extinction learning [4, 82] rather than create a PE experience, due to a change of latent cause. However, the acquisition training memory also expected 100% reinforcement, so the absence of US at CS+ offset would strongly tend to create a PE experience. Inferring which of those competing effects would dominate is beyond current knowledge. Nor can it be inferred whether the next element of the updating procedure, a 10-min waiting period (presumably without video or magazines, as these are not mentioned) prior to extinction training, created a PE experience or induced a change of latent cause, because the authors provide no description of how subjects were led to understand the waiting period or whether SCR equipment was visibly disconnected. If subjects regarded the waiting period as a break, occurring outside of the study, it presumably would not have created a PE experience, in which case the only possible PE experience was created by the absence of the US. If subjects regarded the waiting period occurring within the study, it presumably would have created a PE experience, producing destabilization and annulment even if the absence of the US had not produced a destabilizing PE experience. Thus, PE could have been generated in one or both of those ways, even though reactivation duration was very long.

After the 10-min pause, subjects again viewed the screen and saw a series of unreinforced 6-s lamp-lightings of each color, 8 of each in random order. This extinction procedure functioned as a counter-learning experience while the target learning was reactivated and destabilized.

For a second group of subjects, next on Day 2 after the single, unreinforced, 2-min CS+ reactivation was a 6-hr period before returning to the screen for the extinction procedure. At that point, their target memory had presumably reconsolidated and was no longer in a destabilized condition.

The effects of the Day 2 procedures were tested on Day 3 and Day 5 in different ways. On Day 3 the test was renewal, the unreinforced presentation of a previously extinguished CS in a new context. The new context consisted of lying in an fMRI brain scanner to record brain images of blood oxygen level-dependent (BOLD) activity. SCR data was not collected as subjects viewed 8 unreinforced presentations of each the CSs in headset goggles. On Day 5, subjects were again sitting at the computer screen and a reinstatement test was carried out: First a series of four unsignalled shocks were administered, separated by about 30 s, followed by unreinforced presentation of the lit lamp images of both colors, four times each, with SCR data again being collected. Return of fear was defined as an increase in SCR from the last extinction trial on Day 2 to the first appearance of the shock-associated lamp color after reinstatement on Day 5.

Both the fMRI data on Day 3 and the reinstatement SCR data on Day 5 showed that the 10-min group had no return of fear of the reminded CS+, whereas fear did return for the 6-hr group. After describing in detail their analyses of SCR and fMRI data, the authors conclude that the Day 2 procedure with 10-min interval "abolishes fear expression by erasing a memory trace in the amygdala" (p. 1552).

In a follow-up conducted after 18 months [35], renewal, reacquisition and fMRI tests found that fear annulment had persisted for the 10-min group, but a significant fear response to CS+ was found in the 6-h group, an important confirmation of the enduring nature of the observed annulment of emotional memory.

Relative to the studies by Schiller et al. [6] and Oyarzún et al. [36], Agren et al. [34] achieved annulment using quite different CSs, quite different CS-US reinforcement schedule (100% rather than 38%), and quite different violations of expectation to create PE, but these differences in procedure should not obscure the fact that the same set of three experiences were created: target learning reactivation, concurrent PE experience, and a counter-learning experience.

A.4. Schiller et al. (2013) [38]: "Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement"

This study largely replicated experiment 2 of Schiller et al. [6], with two CS+s and a CS–, with these differences: In acquisition on Day 1, the inter-trial interval was 12 s (instead of 11±1 s), and reactivation on Day 2 consisted of one of the CS+ stimuli presented twice unreinforced, with inter-trial interval of 5 s (instead of one presentation each of a CS+ and the CS–).

Day 2 reactivation by two unreinforced CS+ presentations probably was for some subjects a mismatch that created a PE experience, destabilizing the target memory, because there was low probability of seeing an unreinforced CS+ twice consecutively during acquisition (which contained all three CSs presented 8 times unreinforced and each of the two CS+s presented 5 times reinforced, all in random order). That added source of PE is how the PE analysis of this study differs from that of Schiller et al. [6], given above in Section A.1.

Testing on Day 3 was reinstatement (instead of re-extinction) to measure the persistence or annulment of fear. Annulment of fear was observed for the CS+ that was reminded on Day 2 but not for the unreminded CS+. Thus nullification of the acquired emotional learning resulted from the set of three experiences of target learning reactivation, a concurrent PE experience, and a counter-learning experience.

This study also extended the 2010 study [6] by using fMRI imaging to document the brain regions engaged at all phases of the three-day procedure, a considerably more extensive examination than the fMRI data collected by Agren et al. [34] on Day 3 only. The fMRI data show that the prefrontal cortex was engaged during Day 2 extinction of the non-reminded CS2+ but not the reminded CS1+, adding significant evidence that extinction and reconsolidation are fundamentally different processes.

A.5. Liu et al. (2014) [134]: "An unconditioned stimulus retrieval extinction procedure to prevent the return of fear memory"

These researchers reported a set of six experiments that largely replicated the procedures in experiments 1 and 2 by Schiller et al. [6], including Day 1 acquisition of fear memory using colored squares with a 38% reinforcement schedule and, on Day 2, a 10-min interval between reactivation and extinction. An major variation, however, was reactivating the target fear memory on Day 2 with a single, unsignalled presentation of the US electric shock, with half of the strength used in Day 1 acquisition, rather than with an unreinforced CS+ presentation. In some experiments, another group of subjects had reactivation by a CS+ for comparison.

In one of the experiments, the US-reactivation-extinction procedure was done on Day 15 instead of Day 2 in order to study how the result is affected by a significant increase of memory age.

In another experiment, for acquisition on Day 1, CS1+ was paired with US1 and CS2+ was paired with US2, where the two USs differed in that one was a shock delivered to the right inner wrist and the other was a shock delivered to the right eyelid. Reactivation on Day 2 was done by US1.

The testing of fear recovery or annulment consisted first of re-extinction to detect spontaneous renewal, followed by a reinstatement test. The testing was always done one day after the post-retrieval extinction procedure had been completed. For one experiment, testing was carried out again after 6 months.

The data reported by Liu et al. [134] show these main results:

- Replicating experiment 1 of Schiller et al. [6], but with US reactivation instead of CS+ reactivation, successfully produced updating and annulment of threat memory.
- When two CS+s were paired with the same US on Day 1, Day 2 reactivation by US presentation and extinction by unreinforced presentations of only one of the CS+s resulted in both of those CS+s becoming free of fear at test on Day 3 and also 6 months later. In contrast, Day 2 reactivation by one CS+ presentation resulted in only that one CS+ becoming free of fear. In other words, US-triggered reactivation destabilized all CS+ memory encodings related to that US, whereas reactivation by CS+ destabilized the memory of only that one CS+'s association with the US.

- The same was also shown for an older target memory by carrying out retrieval-extinction on Day 15 instead of Day 2.
- When each of two different USs had been associated with a different CS, US-triggered reactivation-extinction was US-specific, destabilizing and allowing elimination of only the CS+ fear response related to the reactivated US.
- With 24 h instead of 10 min between reactivation and extinction, fear returned at test on the day after extinction, confirming that after 24 h the fear memory was no longer destabilized when extinction was carried out.

The authors do not address the PE requirement or how their various procedures created PE experiences. How the PE evaluation framework of Section 3 applies to US reactivation, showing that multiple PE experiences are almost certainly produced by several salient mismatches with acquisition conditions, is delineated in detail in Section 3.3.2. Briefly, the US occurring alone mismatched the target memory's knowledge from Day 1 that the US always is closely preceded by a CS+. In addition, the significantly reduced (half-strength) shock intensity mismatched the acquisition memory of the intensity. Each of those salient mismatches can reliably be assumed to create a PE experience. Mismatch of the implicit belief that the US is caused by the CS+ is yet another source of PE. For the other experiments that used CS+ reactivation, the PE evaluation of CS+ reactivation after acquisition with partial reinforcement requires a detailed analysis of how the transition from CS+ offset to the "10 min break" was facilitated, as explained in Section 3.3.1. However, Liu et al. [134] provide no information about the transition or what subjects viewed or did during the 10-min period. The fact that the CS+ reactivation experiments did produce updating and annulment of threat memory must therefore be taken to mean, via the PE inference principle (defined in Section 3.2.1), that subjects perceived the 10-min period as being part of the experimental procedure and the unexpected 10-min period therefore created a PE experience that destabilized the reactivated CS+ threat memory.

The various experiments of Liu et al. [134] used a range of different procedures, but in every experiment that produced updating and annulment of threat memory, subjects had the same set of three experiences: target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.6. Steinfurth et al. (2014) [100]: "Young and old Pavlovian fear memories can be modified with extinction training during reconsolidation in humans"

The study by Steinfurth et al. [100] closely replicated experiment 1 of Schiller et al. [6], with these differences: The reinforcement schedule of the CS+ was a slightly higher 50% (8 of 16 CS+ presentations) instead of 38%; testing on the final (third) day of procedures consisted of the reinstatement test (the US presented unsignalled four times, followed later by re-extinction) instead of re-extinction to measure spontaneous recovery; and the entire three-day protocol was carried out both on Days 1, 2, and 3 (applying retrieval-extinction to a one-day-old memory), and also on Days 1, 7, and 8 (applying retrieval-extinction to a 6-day-old memory), the latter version having the purpose of measuring the effectiveness of the retrieval-extinction protocol for destabilizing and nullifying a much older memory, which was this study's particular contribution.

Results were the same for both memory ages: Without reactivation prior to extinction on Day 2 or Day 7, fear (measured via SCR) returned in Day 3 testing with strength comparable to the level at the end of acquisition on Day 1. With reactivation prior to extinction on Day 2 or Day 7, the level of fear in Day 3 testing "was not significantly different from zero" (p. 340). The authors conclude (p. 341), "The present results…suggest that the behavioral interference with the reconsolidation of fear memories could be a useful technique to modify fear memories regardless of their age."

The authors never mention the requirement of PE for inducing memory destabilization, but the same PE analysis applies as was given in Sections 3.3 and A.1 above for the study by Schiller et al. [6]. Thus the study by Steinfurth et al. [100] again demonstrates the same set of three experiences nullifying an acquired fear response: target learning in reactivated state, a concurrent PE experience, and a counter-learning experience.

A.7. Pine et al. (2014) [126]: "Unconscious learning of likes and dislikes is persistent, resilient, and reconsolidates"

The three-day procedure followed by Pine et al. [126] differs significantly from that of Schiller et al. [6], as does the type of emotional memory that was created and then successfully abolished: rather than associative (Pavlovian) fear conditioning formed by primary reinforcement, the memory target was affective preference memory formed by the higher-order learning process of secondary hedonic reinforcement. Furthermore, the acquisition procedure on Day 1 was designed for the target emotional preference to be acquired subliminally, entirely outside of awareness in implicit memory systems. These memory characteristics are representative of a large subset of clinical situations.

The fact that the study by Pine et al. [126] differs greatly in procedure from that of Schiller et al. [6] is reflected in the absence of consideration of the Pine et al. study by any of the retrieval-extinction review articles in the top row of Table 1. Apparently none of the authors of those review articles regarded the Pine et al. study as being in the same category as the retrieval-extinction protocol. That makes the Pine et al. study particularly significant and valuable for the current article's purpose of showing that nullification of an emotional learning requires, according to all available empirical evidence, a set of experiences of certain types, not any particular concrete protocol or procedure used for creating those experiences. Therefore a detailed examination of this study is necessary here.

Acquisition on Day 1 was carried out by Pine et al. [126] through a subliminal instrumental conditioning procedure consisting of repeated, rapid (50 and 67 ms) presentations of two masked visual symbols, one at a time. After each presentation of a subliminal symbol, the subject must respond either "Go" (by pressing the space bar on the computer) or "No-Go" (by not pressing the space bar). For one of the symbols, "Go" wins a small amount of money and the image of a coin appears, and "No-Go" does not win money and the image of a grey rectangle appears. For the other symbol, "Go" loses that same amount of money and "No-Go" avoids that loss. "Subjects were instructed to rely on their gut feeling to make as much money as possible by responding appropriately to the stimuli" (p. 2). The acquisition process had 80 subliminal presentations, 40 of each symbol in random order, and was followed by testing, in which each symbol was presented 20 times without seeing any resulting feedback image, though money was still being won or lost. For each subject on Day 1, that entire acquisition process and testing were repeated with a second, different pair of symbols.

Instrumental learning (operant conditioning) was defined and measured as being a percentage of correct instrumental responses greater than 50% (pure chance), with that correct percentage calculated as the number of "Go" responses following symbols that allow a win plus the number of "No-Go" responses following symbols that could cause a loss, divided by the total number of trials for each subject individually.

Both before and after the entire Day 1 procedure described above, each subject viewed and responded to a set of 60 "discrimination test" presentations designed to detect any conscious recognition of the subliminal symbols. No conscious detection was found.

On Day 2, each subject again went through two rounds of the same learning procedure, but with no testing phase, for 30 presentations of each subliminal symbol in each round. The same masked, subliminal symbols were used as on Day 1, but subjects were not told whether the symbols were the same or different to those in the prior day's trials, nor did subjects know that the contingency was now different: for each symbol, both the "Go" response and the "No-Go" response resulted in a random win or loss outcome image. In the authors' words, the symbols "were now rendered non-differential/discriminatory with respect to their reward and punishment contingencies" (p. 4). This change of contingency was the counter-learning experience being created in Day 2 to see whether it would update the contingency learned on Day 1. Importantly, the authors point out that this change of contingencies "differs from extinction learning where the US is simply omitted, and reversal, where contingencies are entirely switched." (p. 2) The absence of an extinction training in this procedure is perhaps its most significant difference from the retrieval-extinction procedure of Schiller et al. [6], and is presumably the main reason why this study is not considered in any of the review articles in Table 1.

Subjects on Day 2 were assigned to two different conditions, the reconsolidation condition and the control condition, which differed by what preceded the Day 2 learning procedure just described. Subjects in the reconsolidation condition first underwent a "reminder session" consisting of each of the four symbols from Day 1 being presented five times, successively, in the same masked and subliminal manner as on Day 1, and playing for money but without feedback images indicating results, as in the Day 1 testing phases. The 20 presentations in the reminder session had a total duration of "less than 2 min" (p. 4), before which subjects were told that this would be a short test of what they had learned on the previous day and that they would win or lose money in these test presentations. The authors explain, "The purpose of these reminders was to reactivate the learned associations from Day 1 and hence open the hypothetical reconsolidation window." (p. 4) The reminder session was followed by a 10-min period of waiting, after which began the learning procedure described in the previous paragraph.

The Day 2 reminder session of 20 trials, taking less than 2 min and ending in a waiting period, had a duration less than one-third of the typical duration of the several much longer procedures experienced on Day 1 (four of which were 60 or 80 trials, and two of which were 40 trials). It is therefore probable, based on studies of PE created by purely temporal violations of expectation during reactivation [109, 111], that the end of the reminder session coming surprisingly soon was a memory mismatch that created a PE experience, destabilizing the memory of what had been learned on Day 1. What came next, 10 min of unfocused waiting, was certainly an unexpected novelty relative to all of the acquisition experiences on Day 1. The authors explain (p. 4), "Subjects assigned to the reconsolidation condition (the reconsolidation group) underwent a reminder session when they entered the room, waited 10 min and then proceeded to the phase 2 trials." Subjects simply waiting after the reminder trials would be likely to construe the situation as still being within the procedure of the study (in contrast to, for example, subjects of Kindt and Soeter [179] being told they are taking a break as they see researchers removing the electronic equipment that had been attached to them, as reviewed in Section B.3), so this mismatch also presumably induced a PE experience, not a change of latent cause.

The authors of this study deserve ample credit for its various ingenious features. However, they did not recognize the above two probable PE experiences embedded in their procedure, and they wrote, “Our results seem to counter a recent theory that new learning (or the generation of a prediction error) is required during reactivation in order to trigger reconsolidation.…Here, no new learning took place during the reminder…” (p. 11). In response to that statement, Ecker [46] commented (p. 12), “Even researchers who are well aware of the mismatch/prediction error requirement can overlook the occurrence of mismatch in their own procedures.”

Subjects in the control condition had no reminder session on Day 2 and began with the 10-min period of waiting, after which they began the same Day 2 learning procedure as for subjects in the reconsolidation condition. Waiting before beginning any procedures began would not have registered as a violation of expectation of what happens during procedures, i.e., the waiting would not be construed as occurring within and being part of the study, so this initial waiting period in the control condition would not have created a PE experience and therefore would not have destabilized the target memory.

On Day 3, subjects first responded to test presentations identical to those on Day 1: no feedback image of result was provided for each response, and each trial could win or lose money with a “Go” response. There were two rounds of this test, one round for each pair of symbols that was conditioned on Day 1 and rendered non-discriminatory on Day 2, and each round consisted of 80 trials (40 for each symbol, randomized). Data from this testing measured subjects’ levels of instrumentally conditioned responding resulting on Day 3 from the Day 1 and Day 2 procedures.

Next on Day 3 was a recognition test, with each of the symbols that had been subliminal and masked now viewed individually for 3.5 s, in order to detect any conscious recognition and memory of the symbols that were intended to be completely outside of awareness throughout the study. No conscious detection of symbols was found.

Last on Day 3 was a measurement of each subject’s affective preferences for the symbols viewed supraliminally (i.e., in conscious awareness). After receiving the instruction, “Choose the symbol you prefer,” each subject then viewed, sequentially, every possible pair of symbols, drawn from the four symbols used on Days 1 and 2 plus two newly added symbols not used previously. For each displayed pair, the subject chose their preferred symbol, taking as much time as needed for making each choice. Data from this Day 3 test measured each subject’s relative levels of affective preference for the individual symbols. Given that affective preferences “would have presumably been formed during [Day 1] acquisition” (p. 8), this affective preference test on Day 3 was a measure of “the efficacy of the [Day 2] contingency change in learning in altering these associations” (p. 8). In other words, this preference test on Day 3 was designed to answer the question: Were the affective preferences formed on Day 1 updated and unlearned by the Day 2 change of contingency to equality and non-predictiveness of all symbols?

At each subject’s conclusion of all Day 3 testing, the subject was told the total amount of money won or lost in the three days. That amount increased or decreased the pre-agreed fee for participating, and the net payment was made.

These are the authors’ main findings and conclusions:

- Day 1 acquisition training produced a statistically significant level of instrumental response learning based on hedonic preference with no explicit awareness of the symbols driving the learning process. Day 3 tests then showed that Day 1’s instrumental response learning was no longer detectable after the Day 2 training with contingencies changed to be non-differential: “No evidence of conditioned instrumental responses was apparent in either group on Day 3 test trials preceding the preference task. …[T]he percentage of correct responses in test trials on Day 3 did not differ significantly from chance in either group….” (p. 9).
- Quite different results were found for affective symbol preferences. The authors state that “the degree of instrumental learning is not prima facie correlated with the degree of stimulus-outcome learning/preference for each stimulus…” (p. 9). Day 3 measurement of affective symbol preferences showed a large difference between control group subjects and reconsolidation group subjects. When control group subjects chose supraliminally between a subliminal symbol that had allowed a win and a subliminal symbol that had threatened a loss, the former symbol was chosen on 65.5% of all such choices, the latter 34.5%. In sharp contrast, the choices made by reconsolidation group subjects were statistically indistinguishable from chance, i.e., neither category of symbol was chosen more than the other. This means that for reconsolidation group subjects on Day 2, reactivating and destabilizing the Day 1 emotional learnings before carrying out new learning with revised (removed) contingencies allowed that new learning to drive the fundamental unlearning and nullification of the Day 1 emotional preference learnings.

The authors comment, “Lack of persistence (or recovery) of instrumental responses indicates that the conditioned preferences revealed in the preference task were driven by affective properties of the [symbols] (i.e., stimulus-outcome associations) and did not result from any instrumentally conditioned responding.” (p. 9) They add, “the conditioned preferences… exhibited by the control group here are striking, both in their magnitude and their persistence over days following acquisition—enduring the [Day 2] manipulation (which was designed to abolish them) and multiple test rounds.” (p. 10) They conclude, “Our results indicate that despite the strength of appetitive and aversive affective associations, they can also undergo reconsolidation.

…These findings show that reconsolidation is a wider phenomenon than previously described, common to a number of forms of associative learning as well as learning driven by secondary reinforcement such as money, and can occur without awareness.” (p. 11)

Thus, though the study by Pine et al. [126] deployed procedures that differ significantly in several ways from the retrieval-extinction procedure of Schiller et al. [6], the same invariant set of critical experiences is again apparent: target learning in reactivated state, a concurrent PE experience, and a counter-learning experience.

A.8. 2015 Johnson and Casey (2015) [136]: “Extinction during memory reconsolidation blocks recovery of fear in adolescents”

Johnson and Casey [136] carried out a variation of the retrieval-extinction protocol used in experiment 1 by Schiller et al. [6] to test whether fear memory in human adolescents could be eliminated via reconsolidation. That is an age group for whom standard extinction-based therapeutic methods for reducing fear have been found to be less effective than for younger and older age groups [182, 194].

The subjects were 36 adults and 38 adolescents, who were randomly assigned to either the extinction condition or the reconsolidation update condition. The CSs were yellow and blue rectangles viewed on a computer screen, as in the study by Schiller et al. [6], but now each rectangle was a window embedded in a color drawing of a room in a home. The window was black during each inter-trial interval and became yellow or blue for each CS presentation. In acquisition training on Day 1, the room was a bedroom, and on Days 2 and 3, the room was a kitchen. That change of visual context of CSs after Day 1 is another sizable difference in procedure.

In acquisition, the CS+/US pairing was a 50% reinforcement schedule (8 of 16 presentations of the CS+), and the 16 CS– presentations were never paired with the US. All 32 CS presentations were in the bedroom scene. Whereas the US used by Schiller et al. [6] was an electric shock, here the US was a hybrid consisting of an aversive sound presented simultaneously with an aversive image, such as a photo of a snake. Each US presentation had the same sound with a different image. The measure of fear was the skin conductance response (SCR).

On Day 2, subjects in the reconsolidation condition underwent the retrieval-extinction procedure: First they viewed a single unreinforced presentation of the CS+ in the kitchen scene, and then for 10 min a Tom and Jerry cartoon appeared onscreen. That was followed by the extinction procedure, a series of unreinforced, randomized CSs (15 CS+ and 16 CS–) in the kitchen scene. Subjects in the extinction condition began Day 2 by seeing the Tom and Jerry cartoon for 10 min, followed by the extinction procedure (16 CS+ and 16 CS–) in the kitchen scene. Regarding the presentation of the 10-min period itself, the only information provided is (p. 4), “All participants viewed a cartoon video of Tom and Jerry (Warner Brothers) during the 10-minute break, presented on the same computer screen on which they viewed the experiment.”

Johnson and Casey [136], like Schiller et al. [6], never mention PE, though by 2015 the number of studies confirming that PE is required for destabilization had risen to at least 23 (listed in chronological order at https://bit.ly/2b8IbJH). The analysis of PE on Day 2 of this study is as follows, according to the framework in Section 3. Because the target memory did not expect the US to occur with every CS+, the single presentation of the CS+ in the reconsolidation condition did not create a PE experience by being unreinforced. However, the target memory did expect each CS+ (and CS–) to appear in the same visual context as on Day 1, the bedroom scene, as in all 32 CS presentations on Day 1. Seeing the CS+ appear on Day 2 in a different visual context, the kitchen scene, is a salient mismatch and violation of expectations, creating a PE experience for, and destabilizing, the visual and spatial context component of the target memory. It cannot be assumed, however, that the threat memory (CS+/US contingency) component of the memory would also be destabilized by that PE experience, as discussed in Section 3.3.2. Indeed, the possibility that destabilization of the visual component of the target memory might not also destabilize the threat memory is the explanation proposed in Section A.16 below for the return of fear in the reminder condition with altered image of CS+, in the study by Junjiao et al. [117].

A PE experience would also have been created by seeing the unexpected Tom and Jerry cartoon after the reminder CS+ offset, if, and only if, the cartoon and the 10-min break were presented to subjects in such a manner they were regarded as being part of the study, not a “break” occurring outside of the study, as explained in Section 3.3.1. The fact that updating and nullification of the threat memory did result from the reconsolidation condition means that threat memory was destabilized by at least one of those two possible PE experiences.

That analysis could be tested, and the actual PE experience(s) responsible for successful updating could be identified, by reproducing this study in three ways: (a) with the reminder presented in the same visual context as in Day 1 acquisition (bedroom), eliminating the first possible PE experience, (b) without the cartoon and 10-min break, eliminating the second possible PE experience, and (c) with both of those changes, (a) and (b). If either (a) or (b) results in return of fear, the eliminated PE experience was the one that destabilized the threat memory. If neither one of those changes causes return of fear but making both changes does cause return of fear, that would reveal that both PE experiences were created and both contributed to destabilizing the threat memory.

On Day 2, subjects "in the extinction condition started with a 10-minute rest period, in front of the test computer" (p. 4). For these subjects, seeing the cartoon appear was in the context of "resting" prior to resuming participation in the study. Both "resting" in front of the computer and seeing a cartoon on the computer were novel experiences, but neither experience violated the target memory's expectations for what would happen onscreen during the experimental procedure, once it resumed. It had not yet resumed, as they construed the situation subjectively. Therefore no PE was created by this 10-min period for these subjects.

Next on Day 2 for the extinction condition, the extinction procedure of CS+ and CS– presentations began, with the CSs now appearing in the kitchen scene for the first time for this group. That unexpected change of the visual context of CSs could be assumed to create a PE experience, but again the entirety of the procedure has to be considered carefully: The unexpected kitchen context of the CSs occurred in the repetitive series of standard extinction training presentations, so the unexpected kitchen context is exactly analogous to the unexpected absence of the US in standard extinction, which does not create destabilizing PE (apparent from the fact that standard extinction results not in updating but rather a separate extinction memory [16], presumably because standard extinction constitutes a large enough qualitative memory mismatch to induce a change of latent cause [121]). Based on that exact structural analogy, a PE experience was not created by the kitchen context on Day 2, so there was no destabilization for subjects in the extinction condition.

On Day 3, for reinstatement each subject first received a single, unsignalled US presentation, which was followed by re-extinction (16 CS+ and 16 CS–). Return of fear was measured as the difference between the SCR measured in response to the first CS+ on Day 3 and the SCR measured in response to the last CS+ at the end of extinction on Day 2.

The data collected by Johnson and Casey [136] through the procedure described above were characterized by them as follows:

- "Adolescents and adults showed equivalent fear acquisition…" (p. 3)
- "[A]dolescents showed diminished extinction learning over time relative to adults…" (p. 3) In other words, Day 2 extinction of the acquired fear response caused less reduction of that fear response in adolescents as compared to adults (consistent with other research, as noted above).
- "Participants who were reminded of the conditioned stimulus 10 minutes prior to extinction showed no recovery of fear 24 hours later….Conditioned fear did not return in the reconsolidation update group after reinstatement, even in our adolescents who showed diminished within-session extinction [relative to adults], highlighting the robustness of the effect." (p. 3) For both adolescent and adult reconsolidation groups, the level of fear after reinstatement on Day 3 was actually lower than the level of fear at the end of extinction on Day 2, a quite decisive nullification of the acquired fear response.
- In both adolescent and adult extinction groups, after reinstatement the fear response to CS+ was statistically the same as the maximum level generated at the end of acquisition on Day 1. Full-strength fear returned after standard extinction, in sharp contrast to the full elimination of fear in the adolescent and adult reconsolidation groups.
- There were "no significant effects of sex on any phase of the experiment" (p. 3). That aspect of the data appears inconsistent with the strong gender differential found by Chen et al. [133].

For implementing the retrieval-extinction protocol, the procedures used by Johnson and Casey [136] differed significantly from those of Schiller et al. [6], yet the same set of three experiences were created: target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.9. Asthana et al. (2016) [125]: "Preventing the return of fear using reconsolidation update mechanisms depends on the met-allele of the brain derived neurotrophic factor Val66Met polymorphism"

The study by Asthana et al. [125] replicated the design of experiment 1 reported by Schiller et al. [6] except for these differences: The US was auditory, a woman's scream, duration 2 s (instead of electric shock), and during Day 1 acquisition training, the CS+/US reinforcement schedule was a much higher 81%, 13 of 16 CS+ presentations (instead of 38%), significantly increasing subsequent US expectancy with each CS+ presentation. Each CS was onscreen for 4 s, with inter-trial interval of 11±1 s.

On Day 2, subjects in the experimental group first saw a single, unreinforced presentation of CS+ for 4 s, followed by a "10 minutes break", and then an extinction training (16 CS+ and 16 CS–, randomized). Also Asthana et al. [125] make no mention of any onscreen video entertainment during the 10-min period between reactivation and extinction, so it is assumed here that subjects sat quietly without attending to anything in particular during this period.

Day 3 testing for return of fear consisted of a re-extinction procedure (16 CS+ and 16 CS–, randomized) for measuring spontaneous recovery.

The authors acknowledge that prior studies have shown "evidence that a prediction error (i.e. the disparity between learned contingencies during fear acquisition and missing consequences during extinction training) is highly relevant for fear memory modification" (p. 7), though their discussion does not attempt to locate a specific point in their procedure that created the PE experience that must have occurred, given that updating and annulment of emotional memory occurred. For locating that PE experience, a PE evaluation using the framework of Section 3 has to consider that, relative to the PE analysis for the study by Schiller et al. [6] (delineated in Sections 3.3 and A.1), this study created a stronger US expectancy on Day 2 in response to unreinforced CS+ presentations and had no video during the 10-min post-reactivation period.

Even with stronger US expectancy, the unreinforced CS+ reminder would not be a mismatch because during acquisition, 3 of the 16 CS+ presentations were without the US. Therefore, the stronger US expectancy did not create a PE experience, and the only mismatches that could have produced a PE experience were the nonappearance of the expected next CS 11±1 s after the reminder CS+, and the unexpected increased of the 11±1-s interval to 10 min once the next CS did appear.

Those mismatches would actually create PE experiences, however, only if after CS+ offset subjects viewed the 10-min period as being part of the study, as explained in Section 3.3.1. If any communication or behavior by the researchers immediately after CS+ offset had led subjects to believe that the 10-min period was a "break" occurring outside of the study, no PE experiences would develop. Being two sides of the same coin, either each of those two mismatches created a PE experience or neither of them did. The fact that updating and annulment occurred means that both of them did. The 10-min period must have appeared to be part of the study. This is not circular reasoning because it is not an attempt to prove the veracity of this PE evaluation framework, rather it is only a trial run for seeing whether this framework conceivably can account conceptually for all observations. An empirical test of the PE analysis above would consist of exactly replicating the study with the one difference of communicating to each subject at the start of Day 2, "Next you will see one of the images from yesterday presented only once, and right after that will be a 10-minute break that is not part of the study. You can simply relax during that break, and then the study will resume with more images appearing onscreen. It is important for the study that you understand that the 10-min break is not part of the study." Then, during the 10-min period, the computer screen would display the words, "Break time, 10 minutes, not part of the study." Orienting subjects in that way would eliminate the two PE experiences identified above as causing destabilization, which should result in return of fear, by this analysis.

In addition, subjects were genotyped for the functional val66met polymorphism (rs6265) of brain-derived neurotrophic factor (BDNF), in order to examine the possible influence of allelic differences on the reconsolidation process. The authors explain that inconsistent results across reconsolidation studies of fear memory had "led to the proposed hypothesis that allelic differences can moderate the fear reconsolidation mechanism" (p. 7), and they briefly review previous research showing significant influence of BDNF in memory processes. They also reference the one prior human study of reconsolidation with allelic differentiation of subjects [43], which showed a significant difference of results between polymorphisms for two proteins other than BDNF.

The results reported by Asthana et al. [125] added further evidence that allelic differences influence the reconsolidation process. Strength of fear after acquisition on Day 1 was equal between BDNF met allele and non-met allele subjects. For met allele subjects in the Day 2 reactivation (reminder) group, Day 3 re-extinction brought no return of fear, i.e., the threat memory was fully nullified in response to CS+ presentations, whereas the no-reactivation (no-reminder) group showed significant recovery of fear, a distinct indication that reactivation on Day 2 had induced memory destabilization and reconsolidation. In contrast, non-met allele subjects in the reactivation (reminder) group did show significant return of fear on Day 3. Also, for non-met allele subjects there was no significant difference in Day 3 level of fear response between the reactivation and no-reactivation groups, which raises the question of whether destabilization and reconsolidation were induced by Day 2 reactivation for non-met allele subjects.

In short, for met allele carriers the target fear memory was fully nullified by the retrieval-extinction protocol, with all data consistent with a memory reconsolidation process, whereas for non-met allele carriers (carriers of the val66val allele), the fear memory was significantly weakened but not fully nullified, with the totality of data forming an ambiguous picture of the memory processes that occurred for this group.

Therefore, though this study suggests the possibility that for some individuals, emotional memory nullification may require stronger conditions or may not be possible, it nevertheless demonstrates once again that when full nullification of a fear memory does occur, it is the result of the same set of three experiences: target learning in reactivated state, a concurrent PE experience, and a counter-learning experience.

A.10. Golkar et al. (2017) [188]: "Vicarious extinction learning during reconsolidation neutralizes fear memory"

The study by Golkar et al. [188] and that of Thompson and Lipp [74], published simultaneously and described in Section A.11 below, were the first to achieve threat memory updating and annulment through the retrieval-extinction protocol with fear-relevant stimuli used in the acquisition training. The CS+ images of Golkar et al., a spider and a gun, were paired with the US (electric shock) and the neutral, unpaired CS– was the fear-irrelevant image of a mug. Using fear-relevant stimuli increases clinical relevance and, according to Golkar et al., produces "superior conditioning, …typically inferred from their resistance to extinction" (p. 91). Previously the same fear-relevant stimuli had been used by Soeter and Kindt [33] for retrieval-extinction (experiment 2), as well as the same acquisition training design, but the updating and annulment effect was not observed (reviewed in Section B.1).

Another major variation of procedure by Golkar et al. [188] was the use on Day 2 of vicarious extinction, in which the subject viewed a 24-min video showing a computer screen and a person viewing that screen while attached to similar equipment as the subject is, and responding calmly as unreinforced presentations of the study's three CSs appear repeatedly on the screen. Vicarious extinction differs from standard extinction not only by involving additional, more complex processes of learning and memory, but also because, relative to the laboratory context of the Day 1 acquisition training and the Day 3 testing process, the Day 2 safety learning during vicarious extinction occurs in the quite different context of socially witnessing another person's experience of all CSs being shock-free. The authors note (p. 91) that "standard extinction procedures typically yield a highly context-dependent decrease in CR [conditioned response] that recovers when tested in a context different than the extinction context." In effect, therefore, placing Day 2 extinction in a different context than Day 1 acquisition or Day 3 testing asks the question: Is vicarious extinction during the reconsolidation window effective for updating a fear memory acquired in a different context? The clinical relevance of this variation is very great.

Measurement of the fear-potentiated startle (FPS) response as the index of fear and its expression, rather than SCR, is another significant variation in this study. FPS is "a reliable enhancement of the startle reflex when an organism is in a state of fear" (p. 88). Because FPS is driven directly by an amygdala response to threat, it is thought to be more selective and less confoundable than SCR, which is driven by general sympathetic arousal (whether negative or positive affective valence) and by conscious memory and declarative knowledge of the CS-US contingency rather than subcortical amygdala memory, making SCR "highly sensitive to attentional processes" (p. 88). The eyeblink startle reflex in response to a loud noise (40 ms, 104 dB, heard through headphones) was measured by electromyography of the right orbicularis oculi muscle, with peak measured voltage occurring within 200 ms of startle noise onset.

During acquisition on Day 1, each CS (reminded CS+, non-reminded CS+, and CS–) was presented five times for 8 s, with an inter-trial interval that varied between 15, 20, and 25 s with a mean of 20 s. The first presentation of each CS+ was unreinforced, so that the unreinforced first presentation on Day 2 would match the acquisition pattern and therefore not drive any extinction learning. Then followed four consecutive CS+/US pairings. That could be defined as 80% reinforcement, but the chunking of the four reinforced CS+ presentations presumably results in very nearly the same level of US expectancy in response to a subsequent CS+ presentation as a 100% reinforcement. One second before each CS+ or CS– offset (disappearance), the startle probe noise was presented and was followed by the US shock 0.5 s later on reinforced CS+ trials. In addition, the startle noise alone also was presented five times during acquisition in order to measure a baseline startle response without CS+.

For the reconsolidation condition of this study, the Day 2 procedure began with reactivation of the threat memory by a single unreinforced presentation of one CS+ (the reminded CS+). That was followed by a 10-min interval called a "break" during which the subject was offered emotionally neutral magazines to read. That reactivation and post-reactivation procedure also duplicates exactly the study by Soeter and Kindt [33]. Subjects in the non-reminded condition began Day 2 with the magazine break. The break was followed by the vicarious extinction procedure.

Day 3 testing began with reinstatement consisting of three unsignalled shocks. Then, after a 10-min break with magazines, re-extinction began and consisted of 5 non-reinforced presentations of each CS and also the startle noise alone.

Full annulment of fear of the reminded CS+ due to reconsolidation updating is indicated by the Day 3 re-extinction data, in that (a) subjects' FPS responses to the first presentations of the reminded CS+ and the neutral CS– during re-extinction were statistically indistinguishable and (b) the FPS response evoked by the first presentation of the non-reminded CS+ was significantly higher than that of the reminded CS+ and equaled the maximum post-acquisition level of FPS response measured at first presentation on Day 2. In other words, after reinstatement, the reminded CS+ evoked no more fear than the neutral CS–, whereas the non-reminded CS+ evoked fear as strongly as acquisition had generated.

That result indicates that vicarious extinction can be effective as a counter-learning of safety for the reconsolidation updating and annulment of a threat memory. The authors describe this as an "intriguing" result toward "overcoming the contextual dependency of exposure-based treatments" (p. 91). The context-independence of vicarious extinction of a destabilized memory is yet another fundamental difference from the phenomenon known as "extinction," again supporting the view that the term "extinction" used for MR updating procedures is always a misnomer.

The fact that updating and nullification of the threat memory did result from the reconsolidation condition means that at least one PE experience was created (though PE is not addressed by the authors). Locating the creation of a PE experience begins, as always, with closely studying the acquisition training as a map of the contents of the target memory. It is probable but not certain that, because the chunked structure of the Day 1 acquisition training functioned in effect as 100% reinforcement, the Day 2 unreinforced CS+ presentation functioned as a mismatch and created a PE experience. It is also probable but not certain that encountering a 10-min break with magazines instead of the expected continuing series of CS presentations every 20±5 s did not create a PE experience because the offering of magazines to read is a strong indication to subjects that this 10-min period is not part of the study, is not happening due to the same latent cause as the Day 1 procedure and the Day 2 CS+ presentation, and therefore is not a violation of expectations based on that prior latent cause, as explained in Section 3.3.1. Thus, the PE experience must have been created by the unreinforced CS+ reminder presentation. That PE analysis could be tested by replicating the study exactly except for the acquisition training consisting of 50% reinforcement, with US pairing on every other reminded CS+, and with subjects guided to recognize explicitly that reinforcement pattern. Then the single unreinforced CS+ reminder presentation would not create a PE experience, threat memory would not be destabilized, and fear would return in Day 3 testing, if this PE analysis is correct.

This study's procedure differed significantly from that of other successful annulment studies, but no variation is seen in the set of experiences created by the procedure: target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.11. Thompson and Lipp (2017) [74]: "Extinction during reconsolidation eliminates recovery of fear conditioned to fear-irrelevant and fear-relevant stimuli"

As in the study by Liu et al. [134] described in Section A.5 above, Thompson and Lipp [74] again used US-reactivation to destabilize and then nullify multiple conditioned fear associations, and they extended that procedure to fear-relevant CSs (the significance of which was noted above in Section A.10 in describing the study by Golkar et al. [188]). The fear-relevant images were a spider and a snake (CSa+ and CSa-), and the fear-irrelevant images were a blue and yellow square (CSb+ and CSb-). The two CS+s were always paired with an electric shock during Day 1 acquisition (100% reinforcement) and the two CS-s were never paired with shock. Each CS was visible for 6 s and each of the four CSs was presented a total of 8 times, in randomized order, with an inter-trial interval of 12±2 s. SCR measurements served as the index of fear response.

On Day 2, subjects in the US-reactivation group first received one unsignalled shock at half the intensity used on Day 1, also as in the study by Liu et al. [134]. After the US reactivation, subjects were given a 10-min break with magazines available for reading. Subjects in the no-reactivation control group began on Day 2 with the 10-min magazine break. For both groups, the break was followed immediately by an extinction training consisting of 10 unreinforced, randomized presentations of all CSs. At the end of extinction, all CSs were receiving statistically indistinguishable, low-level SCR responses, in both groups.

Testing on Day 3 began with re-extinction (8 unreinforced, randomized presentations of each CS) for measuring spontaneous recovery. Then, after providing subjective ratings of CS pleasantness vs unpleasantness, subjects received three unsignalled shocks, 6 s apart, at the same intensity used during acquisition, for reinstatement. A 10-min break with magazines came next, and then 8 unreinforced, randomized presentations of each CS for measuring the effect of reinstatement.

Day 3 test results show that for the US-reactivation group, in response to both re-extinction and reinstatement there was no differential return of fear, i.e., statistically the same SCR levels for CSa+/- and for CSb+/-, whereas the no-reactivation control group, which underwent a standard extinction procedure, did have significant differential return of fear in both tests.. Thus threat memory of both fear-relevant and fear-irrelevant stimuli were abolished for the US-reactivation group. This is the first demonstration of annulment of threat memory of a fear-relevant stimulus through US reactivation. It is also replication of the demonstration by Liu et al. [134] that reactivation by the US destabilizes the threat memory of each CS+ associated with that US in humans, allowing updating of each threat memory by safety learning that directly contradicts, revises and nullifies each of those target memories simultaneously (as was first observed in animals studies [95]).

The authors comment on the creation of PE by the US appearing alone, without being preceded by a CS+ as in the acquisition training. (It is noteworthy that this is the first peer-reviewed journal article reporting successful reconsolidation annulment of a human emotional memory in which the authors provide a detailed consideration of the PE requirement.) Additional PE experiences almost certainly were created simultaneously by the reminder shock intensity being significantly weaker than on Day 1 and also by seeing that the cause of the US appearing is not the CS+ appearing, that latter belief having been formed in the acquisition training. (See Section 3.3.2 for a fuller account of how the PE evaluation framework of Section 3 applies to US reactivation.)

This study again shows that the use of diverse procedures for behavioral updating and annulment does not change the creation of the same three experiences of target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.12. Li et al. (2017) [113]: "Moderate partially reduplicated conditioned stimuli as retrieval cue can increase effect on preventing relapse of fear to compound stimuli"

Use of a compound, 3-element image for CS+ and a different 3-element image for CS– allowed Li et al. [113] to present either 1, 2, or all 3 of the elements of CS+ as a retrieval cue on Day 2. Each visual element was a uniquely colored geometrical shape depicted as three-dimensional. Those were this study's main variations on experiment 1 of Schiller et al. [6].

Day 1 acquisition training had other, minor variations, namely, 20 presentations of CS+ and CS- for 5 s each and an inter-trial interval of 9±1 s, with 12 of the CS+ paired with an electric shock US for 60% reinforcement. Fear responses were measured as SCR.

On Day 2, an unreinforced retrieval cue was presented for 5 s, consisting of either 1/3, 2/3, or 3/3 of the 3-element CS+ image used during acquisition. This was followed immediately by a 10-min "break" during which subjects watched an excerpt from the BBC nature documentary Planet Earth. Immediately after that viewing, extinction training began and consisted of 8 unreinforced presentations of CS+ and 9 of CS−. For a fourth group of subjects, no retrieval cue was presented.

Day 3 testing began with re-extinction (8 unreinforced, randomized presentations of each CS) for measuring spontaneous recovery. Reinstatement began 18 s after re-extinction with four unsignalled shocks, 1 s apart. After a 5-min break, 9 unreinforced, randomized presentations of each CS served to measure the effect of reinstatement. Data from those tests revealed the following responses of the CS+ fear memory:

- 3/3 reminder: strong return of fear in spontaneous recovery, no return of fear after reinstatement;
- 2/3 reminder: no return of fear in spontaneous recovery or after reinstatement;
- 1/3 reminder: strong return of fear in spontaneous recovery and after reinstatement;
- no reminder: strong return of fear in spontaneous recovery and after reinstatement.

Li et al. [113] acknowledge the critical role of PE for memory updating with a few general comments. The PE evaluation of this study according to the framework in Section 3, accounting in detail for the diverse findings listed above, is as follows.

The 3/3 reminder is the exact CS+ presented during the partial reinforcement acquisition training, so it had zero mismatch of both the CS+ visual image and the CS+/US contingency components of the CS+ memory. The absence of the US also was not a mismatch of the partial reinforcement during acquisition. However, the unexpected singleness of the CS+ reminder, due to the unexpected appearance of a video onscreen, was a mismatch of the expected long series of CSs that occurred in acquisition. The unexpected video itself also was a salient mismatch.

That entire reactivation and post-reactivation description is identical to that of the studies of Schiller et al. [6] experiment 1 and Oyarzún et al. [36] (reviewed in Sections A.1 and A.2, respectively). However, both of those studies found that Day 3 re-extinction brought no return of fear, whereas it did re-evoke fear as reported by Li et al. [113] for the 3/3 reminder. Clearly, the fact that fear could be re-evoked by Day 3 CS+ presentation means that the threat memory was less fully updated and annulled by Li et al. using the 3/3 reminder. Explaining that difference is necessary for understanding this study.

A comparison of experimental parameters reveals that, relative to studies by Schiller et al. [6] and Oyarzún et al. [36], Li et al. [113] created a significantly stronger threat memory on Day 1 and used a significantly weaker counter-learning on Day 2 that presumably failed to update the threat memory decisively. For acquisition training, the comparison of reinforcements is 60% versus 38%, the comparison of number of acquisition trials is 20 versus 16, and the comparison of number of acquisition CS+/US pairings is 12 versus 6. The acquisition training carried out by Li et al. had twice as many CS+/US pairings as presented in acquisition by Schiller et al. experiment 1 and Oyarzún et al. Thus, Li et al. created a target memory having significantly stronger threat due to stronger memory of suffering (double the number of CS+/US pairings) and stronger US expectancy and certainty of latent cause (due to higher rate of reinforcement).

The post-reactivation extinction training of Li et al. [113] had 8 CS+ unreinforced presentations compared to 12 CS+/US pairings in acquisition, a ratio of 0.67. The extinction training of Schiller et al. [6] and Oyarzún et al. [36] had 10 CS+ presentations compared to 6 CS+/US pairings in acquisition, a ratio of 1.67. Thus Li et al. had 0.67/1.67 = 40% of the counter-learning weight as in the other two studies, for a target memory that was about twice as strong, making the net effective counter-learning weight about 20% in the study by Li et al. relative to the other two studies. Weak counter-learning is uncertain counter-learning. Allowing an uncertain counter-learning to nullify an existing emotional memory would not be a survival-positive evolutionary adaptation. It seems plausible that a counter-learning weakened to 20% could fail to update a threat memory well enough to prevent all return of fear, but it still remains to explain why on Day 3 it was the CS+ presentation and not the US presentation that could evoke return of fear after reactivation with the 3/3 reminder.

That question may be answerable by asking a different question: What induces fear more strongly, another stab of the pain felt during a recent mugging, or a fresh encounter with the mugger? The US on Day 3 is another stab of the recent pain, and the CS+ on Day 3 is a fresh encounter with the entity that appeared to be inflicting that pain during acquisition on Day 1. A technical analysis of that memory phenomenology would have to include the episodic memory of Day 1, i.e., memory of the experience of repeatedly seeing the CS+ appear again and knowing it could inflict pain in a few seconds and feeling helpless to do anything self-protective. Thus for a human being who paid attention to the Day 1 acquisition training as instructed, on Day 3 the CS+ evokes a subcortical fear response significantly more strongly than the US does. Evidently, after the 3/3 reminder, the Day 2 counter-learning used by Li et al. [113] was sufficient to prevent return of fear in response to the US, but not sufficient to prevent return of fear in response to the significantly more fear-provoking CS+ on Day 3.

In the 2/3 reminder condition, mismatch of the visual image of CS+ was introduced, subjecting the CS+ visual memory to a PE experience, which additionally destabilized the visual image component of the CS+ fear memory. This is an example of the PE specificity phenomenology introduced in Section 3.3.2. The results with the 2/3 reminder (listed above) show that more of the CS+ threat memory was then updated and nullified than with the 3/3 reminder, eliminating fear responses to CS+ in re-extinction on Day 3. With the CS+ memory itself destabilized, the subsequent Day 2 extinction procedure presumably connected the CS+ image memory directly to a revised encoding of the CS+/US contingency, namely the knowledge that the CS+ appears and disappears without the US happening. Then the CS+ presentations on Day 3 evoked no fear. (Study authors Li et al. [113] interpret the differential effect of the 2/3 reminder, as compared with the effect of the 3/3 reminder, in terms of the visual salience of the mismatching 2/3 reminder inducing activation of the locus coeruleus/norepinephrine (LC–NE) system, which is known to respond to a salient or behaviorally significant stimulus. That effect would augment the PE phenomenology proposed here.)

The 1/3 reminder is this study's largest degree of visual mismatch relative to acquisition conditions. In sharp contrast with the 2/3 reminder case, fear returned strongly in response to both CS+ presentation (spontaneous recovery test) and US presentation (reinstatement test), which is identical to the result found in the no reminder case consisting of standard extinction training. That result indicates that the 1/3 reminder mismatch was too large to create a PE experience and instead induced a change of latent cause, so no destabilization or updating occurred and the CS+ fear memory was only temporarily suppressed by standard extinction. This corresponds to the super-salient mismatch in the right-hand section of Figure 1A.

The PE specificity model introduces the concept of component-specific, destabilization, updating and annulment of memory, with subsequent differential cue-specific responsiveness of the memory, as is apparent in comparing the test results of the 2/3 and 3/3 reminders above. The different effects of the 1/3, 2/3, and 3/3 reminders are not adequately described by saying that the reminders constituted different degrees of mismatch and PE. The 2/3 and 3/3 reminders had different effects because they mismatched different components of the target threat memory, according to the PE specificity analysis.

In the 2/3 and 3/3 reminder conditions that produced updating and annulment of fear responding, it is apparent that the procedure created the three experiences of target memory reactivation, a concurrent PE experience, and a counter-learning experience.

A.13. Hu et al. (2018) [150]: "Reminder duration determines threat memory modification in humans"

Using essentially the same protocol as in experiment 1 of Schiller et al. [6], fear memory acquisition on Day 1 consisted of one CS+ (a colored square) paired with shock US at 38% reinforcement, the CS– (a different-colored square) was never paired with US, each CS was onscreen for 4 s with inter-trial interval of 11±1 s, and fear was measured via SCR. Hu et al. [150] expanded the 2010 study [6] by carrying out the protocol for five groups of subjects using five different unreinforced reminder CS+ durations on Day 2: durations of 1 s, 4 s, 30 s, 3 min, and no reminder.

Two other variations from the original version of the protocol by Schiller et al. [6] are also noteworthy: For the 10-min break following the Day 2 reminder, subjects were not shown a video, rather they were instructed to rest. The Day 3 tests added reinstatement following spontaneous recovery.

For the 1-s and 4-s reminder groups, Day 3 data show annulment of the acquired fear memory for CS+, indicated by no statistical difference between SCR levels for CS+ and CS– in both the spontaneous recovery and re-extinction tests. In contrast, fear memory recovered significantly in Day 3 tests of the no reminder, 30 s and 3 min groups.

Hu et al. [150] cite previous studies that found the same pattern: shorter reminders destabilize the target memory, allowing its annulment, whereas long-duration reminders do not [4, 82]. They discuss how their results can be understood in terms of various possible explanatory principles, namely prediction error (PE), the trace dominance principle [183], and latent causes [121](. However, their discussion is in general terms and does not examine their own specific procedure to locate PE creation. The possible roles of trace dominance or latent causes in the Hu et al. study become apparent in the course of applying the framework of PE analysis mapped in Section 3, as shown below.

First, the 1-s and 4-s durations that produced annulment: In acquisition training on Day 1, subjects learned that a CS square always remains visible onscreen for 4 s, and that CS+ is sometimes, but not always, accompanied by a shock, as noted above. On Day 2, amygdalar reactivation of the CS+ fear memory presumably occurred within 300 msec of CS+

appearing onscreen [192]. CS+ offset (disappearance) after 1 s is distinctly shorter than the expected 4-s duration, creating a temporal mismatch and PE experience that can be assumed to have induced destabilization [109, 111].

A 4-s presentation of CS+ exactly matched the expected duration, so no PE was created by CS+ offset, and absence of the US also did not violate expectations due to 38% partial reinforcement during acquisition training. The expectation then was for an inter-trial interval of 11±1 s, but the next CS image appeared after 10 min, a salient mismatch of that expectation. That mismatch is the only possible source of the PE experience that must have occurred, given that updating and annulment was the result. That implies in turn that subjects construed the rest period as being a part of the study, preserving the active latent cause, otherwise a PE experience would not have occurred (as explained in Section 3.3.1).

The 30-s and 3-min reactivation presentations of CS+ on Day 2 did not result in memory updating, which implies that no PE in relation to the CS+ memory was generated by these durations. The absence of PE can be understood as being due to a change of latent cause: In relation to the expected 4-s duration of CS+, a 30-s duration was presumably a mismatch too large for the CS+ memory's original latent cause to remain relevant and operative (right-hand portion of Figure 1A). That is, CS+ with 30-s duration was construed by the subject's emotional learning system as being a different category of phenomenon with a different latent cause, i.e., produced by a different state of the world, than the CS+ in the original target memory. Therefore the original CS+ memory ceased to function as the relevant, active context for predicting or making sense of current experiencing at some point in time prior to CS+ offset at 30 s. Consequently the duration mismatch of 30 s versus the expected 4 s did not function as PE in relation to the now contextually disengaged CS+ fear memory. Undoubtedly, being asked to rest for 10 min after CS+ offset at 30 s or 3 min was unexpected and surprising for subjects, but that surprise would not create PE in relation to the disengaged CS+ fear memory, according to this latent cause analysis.

Thus, this analysis can explain why a PE experience was created in the 1-s and 4-s reminder conditions of Hu et al. [150], each of which resulted in CS+ memory annulment, and why a PE experience was not created in the 30-s, 3-min, and no reminder conditions, which resulted in return of fear. That account also maintains the gleaning that a set of three experiences produces emotional memory annulment: target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.14. Chen et al. (2018) [180]: "Effects of prediction error on post-retrieval extinction of fear to compound stimuli"

The three-day retrieval-extinction procedure of Chen et al. [180] elaborated the procedure used by Sevenster et al. [115, 116] for systematically varying the amount and types of mismatch present during reactivation on Day 2, in order to study the effects of different amounts and types of PE in relation memory destabilization and updating. A Day 1 acquisition training with reinforcement occurring uniformly on every other CS+ presentation allows for Day 2 reactivations in various patterns that have well-defined amounts and types of mismatch. The results obtained by Sevenster et al. contributed importantly to researchers' recognizing that PE is critically required for inducing destabilization that allowed pharmacological disruption of the reconsolidation process to abolish the memory.

Acquisition training consisted 6 CS+ and 6 CS–, with 3 CS+ accompanied by electric shock US, alternating with the 3 unreinforced CS+. The skin conductance response (SCR) was the measure of fear. With that training, subjects would be expecting that same alternating pattern of reinforcement on Day 2. These were 8-s presentations with 9±1 s inter-trial interval, with "Please rest" displayed onscreen during the inter-trial period. Each CS was a compound (two onscreen geometrical shapes and a neutral sound, all simultaneous), but there was no varying of CS composition. What was varied was the Day 2 reactivation presentation, a different mismatch for each of four groups of subjects:

- one unreinforced CS+ ("no PE group");
- two unreinforced CS+ ("negative PE group");
- two reinforced CS+ ("positive PE group");
- four unreinforced CS+ ("multiple PE group").

Following reactivation, subjects "took a rest for 10 min" and then saw a randomized, unreinforced presentation of 10 CS+ and 10 CS– again with 9±1 s inter-trial interval.

Day 3 memory testing began with re-extinction (8 CS+, 8 CS–) to measure spontaneous recovery. After 1 min, four unsignalled shocks, 1 s apart, began a reinstatement test, and 5 min later an unreinforced series of 7 CS+ and 7 CS– was presented.

There was no return of fear for "negative PE" and "positive PE". Each of these reactivations is a definite, salient mismatch, i.e., a violation of the threat memory's expectation that the shock will occur on every other CS+, exemplifying the middle section of Figure 1A.

There was strong return of fear for "no PE" and "multiple PE". One unreinforced CS+ is not a mismatch, corresponding in Figure 1A to a mismatch of zero. Four unreinforced CS+ is a series of two mismatches of the acquisition memory's expectation, double the mismatch of the "negative PE group". The fact that this did not result in memory annulment means, according to the PE evaluation framework in Section 3.3.1, that subjects perceived a series of two distinct mismatches as being too large to be relevant to the acquisition memory or to be caused by same state of world as caused the acquisition training. For one mismatch, the threat memory's expectation remains relevant but it erroneous, whereas for two mismatches, the threat memory's expectation is not relevant and therefore is not erroneous. A different latent cause was induced, so no PE experience and no destabilization occurred, corresponding to the right-hand section of Figure 1A. In this conceptualization, which differs from that of Chen et al. [180], reactivation by four unreinforced CS+ is multiple mismatch but not multiple PE or excessively large PE, because a mismatch that is too large to maintain relevance to the target memory does not create a PE experience at all, rather it induces a change of latent cause.

Chen et al. [180] have contributed an efficient demonstration of the full range of mismatch and PE phenomenology, including a reiteration of the invariant set of experiences that result in emotional memory annulment, namely the reactivated state of the memory, an unambiguous PE experience in response to a mismatch that is relevant to the memory, and then counter-learning.

A.15. Grégoire and Greening (2019) [85]: "Opening the reconsolidation window using the mind's eye: Extinction training during reconsolidation disrupts fear memory expression following mental imagery reactivation"

Grégoire and Greening [85] used visual CSs and an electric shock US and carried out the same overall procedure as in experiment 2 of Schiller et al. [6], with acquisition on Day 1 with 38% partial reinforcement of two CS+s and no reinforcement of the CS–. The reminded CS+ and the non-reminded CS+ were each presented 13 times, 5 of those with US, and the CS– was presented 8 times, all randomized and with inter-trial interval of 13±1 s. The CS images were different orientations of the Gabor patch, a soft-striped graphical element used in memory and perception research. Skin conductance response (SCR) was measured to detect fear and reinstatement (unsignalled US presentation) tested for return of fear on Day 3.

An important variation was the reactivation procedure on Day 2: Rather than reactivating the target memory by presenting one of the two CS+ images once as an external perception, each subject in the reminded condition instead was guided to visualize an unreinforced CS+ internally, as vividly as possible, for 4 sec. The aim of Grégoire and Greening [85] was to see whether mental imagery could reactivate a memory and induce its destabilization, allowing counter-learning to then nullify it. That visualization was followed by a 10-min "break" during which a BBC nature video was playing onscreen, after which extinction training began (10 reminded CS+, 11 non-reminded CS+, 11 CS–, all randomized and unreinforced).

Day 3 memory testing consisted of reinstatement by four unsignalled shocks, followed by another 10-min break with video, and then re-extinction for measuring fear responding.

The authors summarized the test data in this way (p. 279): "skin conductance reactions, indexing fear, did not recover for the reminded (mentally imaged) CS+, whereas it did recover for the non-reminded CS+. This attests that mental imagery facilitated reconsolidation of a conditioned fear association. … To our knowledge, this study is the first demonstration of fear attenuation using imaginal reactivation." That success in abolishing the fear response after imaginal reactivation is important in supporting the use of imaginal techniques in the therapeutic recruitment of MR.

The matter of PE is not addressed by Grégoire and Greening [85]. The fact that updating and annulment occurred means that the Day 2 imaginal reactivation procedure did create at least one PE experience, destabilizing the target CS+ memory. Discerning exactly in the procedure where that PE occurred requires comparing a detailed account of the Day 1 acquisition experiences with a detailed account of the Day 2 reactivation experiences, as follows. Each of the 34 acquisition trials consisted of this set of perceptions:

- black dot appears at center of computer screen for 2 s
- black dot remains as one of the three CSs appears onscreen for 4 s
- 5 of the 13 presentations of each CS+ end with electric shock
- black dot turns white at CS offset for inter-trial period of 13±1 s

Each of those percepts has to be regarded as a component of the target memory.

The Day 2 procedure began in this way: "Before the start of the reactivation phase, participants were informed that this trial was an imaginal trial involving one of three stimuli from the previous day." (Supplemental materials, p. 6) Subjects then viewed the computer screen and saw the familiar black dot appear for two seconds, then they visualized the reminded CS+ for 4 s, and then the black dot turned white for 13±1 s. Because the visualizing experience was done within the same familiar sequence and context as the Day 1 acquisition training, subjectively it had the same latent cause as Day 1 acquisition training, so the mismatch of seeing the reminded CS+ internally rather than externally created a PE experience. This PE experience (i.e., the unexpected encounter with the CS+ internally) implicitly raises the question in each subject as to whether the CS+/US contingency follows the CS+ into the inner imaginal domain. Due to its relevance to the CS+/US contingency, this PE experience destabilizes the threat memory because the threat memory now needs updating regarding its relationship to the imaginal domain.

There is also the possibility that the next event, an unexpected 10-min break with a nature video on view, might also have created a PE experience. Again all details must be considered to evaluate this possibility. Fortunately, the authors have provided the needed information (p. 278): “During all experimental stages, except for the habituation phase and the breaks, the participants were attached to the SCR and shock electrodes, and the shock stimulator was set to the ‘On’ position.” The disconnecting of the SCR and shock electrodes for the 10-min break certainly led subjects to regard the break and the video as occurring outside of the study with a different latent cause, so no PE experience could have occurred. The only possible PE experience in the study, by this analysis, is the one due to the internal, imaginal viewing of the reminded CS+, described above.

With the target memory of the reminded (imagined) CS+ destabilized by that PE experience, the subsequent 32 extinction trials on Day 2 functioned as an unlearning and updating experience, annulling the target threat memory by replacing it with the knowledge and expectation that the reminded CS+ is a harmless perception. Reinstatement on Day 3 therefore did not recover fear of the imaginally reminded CS+.

Though the imaginal procedure used for memory reactivation by Grégoire and Greening [85] was a major variation, nothing was different in terms of types of experiences created to produce emotional memory annulment, namely target memory reactivation, a target memory PE experience concurrent with reactivation, and a counter-learning experience soon thereafter.

A.16. Junjiao et al. (2019) [117]: “Role of prediction error in destabilizing fear memories in retrieval extinction and its neural mechanisms”

Junjiao et al. [117] combined the use of compound CSs with the procedure used by Sevenster et al. [115, 116] for systematically varying the amount of mismatch present during reactivation on Day 2, in order to study the role of PE in causing memory destabilization.

Junjiao et al. [117] aimed to learn whether PE and destabilization are necessary also for the behavioral retrieval-extinction updating process, as they had been shown to be for pharmacological annulment of emotional memory. In addition, they aimed to find whether destabilization strictly requires PE or whether a stimulus novelty that does not create PE can also induce destabilization.

In addition to measuring SCR as an index of fear responding, fMRI images of brain response were also collected on Day 2 and Day 3 in a brain imaging center, with each subject lying on the fMRI scanning bed and the computer screen mounted so as to be easily visible. The fMRI data was sought in order to study the neural correlates of the memory processes operating.

Acquisition training on Day 1 was done in a different context, a behavioral laboratory. The CS+ and CS– each consisted of three perceptual components: two distinct visual, geometric elements and one sound, all presented together. (Previously, Li et al. [113] had also used compound, three-element CSs, reviewed in Section A.12 above.) The CS+ and CS– were each presented 10 times, randomized, and every other CS+ presentation was paired with an electric shock US, for 50% reinforcement schedule. Subjects were guided before and after acquisition to also acquire explicit declarative knowledge of the rule that the shocks occur regularly on every other CS+ presentation. Each CS appeared for 8 s, and during each inter-trial interval of 9±1 s, the words “Please relax” were displayed onscreen.

On Day 2, subjects in three different groups were presented with different unreinforced 8-s reminder/reactivations in the fMRI unit:

1. two of the three components of the CS+ (either one of the visual elements with the sound or two visual elements without the sound) presented once;
2. the complete, three-element CS+ presented once;
3. the complete, three-element CS+ presented twice.

Reminder offset was followed by the words “Please relax” displayed onscreen for the same inter-trial interval of 9±1 s as during acquisition, and then these words were displayed onscreen: “Now please rest 10 minutes. Please open your eyes and stay awake.” After 10 min, extinction training began and consisted of 12 unreinforced, randomized presentations of CS+ and CS–.

Return of fear of CS+ was tested on Day 3 with a re-extinction procedure (12 each of unreinforced, randomized CS+ and CS–) to measure spontaneous recovery as the difference between SCR level on the first Day 3 re-extinction trial and the last Day 2 extinction trial.

The test data show that there was no return of fear of CS+ only for reminder #3 above. That complete, three-element, unreinforced CS+ presented twice was a salient violation of expectation of the well-defined CS+/US contingency learned in acquisition, creating a PE experience in exactly the same way as in the studies by Sevenster et al. [115, 116].

Reminder #2, the complete, three-element, unreinforced CS+ presented once, was consistent with the CS+/US contingency learned in acquisition, so this condition is termed the no-PE group by Junjiao et al. [117]. CS+ fear returned significantly for this group.

Reminder #1, with two of the three components of the CS+ presented once, also resulted in significant return of CS+ fear, and is conceptualized by Junjiao et al. [117] as not creating a PE experience because it presents a novelty that has no relevance to the CS+/US contingency, though it has enough of the original three-part CS+ image to definitely reactivate the target memory. Another result of reminder #1 is a significant SCR level in response to CS– on Day 3, equal to the SCR level for CS+.

The effects of reminder #1, including the SCR response to CS–, can be explained by the PE evaluation framework of Section 3 as follows. In this analysis, the two-part CS+ discrepancy did create a PE experience specific to the CS+ sensory representation components (visual and auditory) in the original acquisition memory, destabilizing that sensory representation but not the CS+/US contingency (threat) representation. That would be the PE specificity effect discussed in Section 3.3.2, and it would be consistent with the cue-specific destabilization of specific memory components demonstrated by several studies [6, 36, 38, 128, 134]. With only the encoding of the CS+ sensory representation destabilized, the subsequent extinction procedure did not update the threat memory, rather the presentations of the CS– caused new encoding of the CS– sensory representation to become conjoined or integrated with the destabilized CS+ sensory representation, which still had its own linkage to the CS+/US contingency (threat) memory. The CS– presentations during extinction training did not update the CS+ sensory representation because the extinction procedure also freshly presented and maintained the same original CS+ sensory representation as during acquisition. The new compound CS+/CS– encoding with its threat memory linkage was then the memory that responded during Day 3 re-extinction to both the CS+ and CS– presentations, which is why the SCR amplitude was the same for both CS+ and CS– presentations on Day 3. In contrast, with the other two reminders, the separate CS+ and CS– memories formed during acquisition still were the memories that responded, separately, to the CS+ and CS– images presented during Day 3 re-extinction.

Heightened SCR amplitude in response to CS– has been reported in other human retrieval-extinction studies without any change made in the reminder cue, e.g. [50, 184], as well as in extinction studies (for references, see [50]). A number of possible causes of the effect have been discussed [50], including unintended contextual fear conditioning during acquisition, causing any subsequent stimulus presentation to induce a heightened emotional response. The analysis proposed above for the CS– SCR level reported by Junjiao et al. [117] for reminder #1 makes unique predictions that could be tested. For example, it predicts that if Day 2 extinction following reminder #2 were to consist of only CS+ presentations and no CS– presentations, the CS– image memory would remain separate from the CS+ image memory (as for the other two reminders), so SCR amplitude for CS– would remain significantly lower than for CS+ in Day 3 re-extinction.

The fact that reminders #1 and #2 resulted in return of fear, indicating no destabilization of the CS+ threat memory, means that the 10-min rest period, which was the same in all three conditions, did not itself create a CS+/US contingency PE experience and therefore did not destabilize the threat memory. Explaining why the 10-min period did not create PE in this study is required of the PE evaluation framework of Section 3.

In all three conditions, reminder offset was followed by the expected inter-trial presentation learned during acquisition, namely the words "Please rest" displayed for 8 s. After reminder #2, which was the complete, three-element CS+ presented once, the target memory's expectation was now for another CS image to appear, but what appeared instead was a novel display of words: "Now please rest 10 minutes. Please open your eyes and stay awake." A PE experience would certainly be created by that violation of expectation. The challenge here is to understand why that PE experience was specific to a memory component other than the CS+/US contingency and therefore did not destabilize the threat memory. The expectation that a CS+/US pairing would come next after reminder #2 may have been weakened by two different features of the procedure: the change of context from acquisition (sitting upright in a chair in a laboratory) to reactivation (lying on the bed of an fMRI machine), and the fact that the acquisition sequence was a randomized mixture of CS+s and CS–s, so the number of CS presentations between CS+/US pairings was unpredictable and variable up to 5 (CS+US / CS¬– / CS– / CS+ / CS¬– / CS– / CS+US). Presumably, after one unreinforced CS+ reminder, the uncertainty of US presentation was high enough (i.e., US expectancy was low enough) due to those two effects that cutting off the expected ensuing series of CSs was sub-salient mismatch of specifically the CS+/US contingency (see left-hand section of Figure 1A), and therefore did not create the PE that would destabilize the threat component of the total acquisition memory, though presumably it did mismatch and create PE that destabilized the more general semantic memory of CSs always appearing in a series. Reminder #1, being discrepant from the CS+ visual image memory, introduced even more uncertainty that further decreased US expectancy.

The study's aim of clarifying the role of PE in the retrieval-extinction procedure was achieved by demonstrating that PE is necessary for destabilization that allows behavioral updating and nullification, as it was already known to be for pharmacological abolition of a target learning. Based on that finding, the authors state, "retrieval extinction…targets the direct rewriting of the CS-US connection…" (p. 303).

In summary, annulment of a learned fear was accomplished by Junjiao et al. [117] using new procedures for acquisition and reactivation, but no change in the set of three experiences consisting of target memory reactivation, a PE experience specifically relevant to the target fear memory, and a counter-learning experience soon thereafter.

The authors also present extensive fMRI evidence allowing them to state that "the neural mechanisms assessed through brain imaging that differentiate effective retrieval extinction from standard extinction are that the former diminishes the IT cortex and prefrontal cortex (particularly, dlPFC) involvement and functional connectivity of IT-dlPFC and dlPFC-ACC…" (p. 304). "We argue that the [greatly reduced] involvement of the dlPFC in Day 2 extinction can distinguish the two paradigms." (p. 303) These findings add considerable weight to the conclusion from several earlier studies that the memory reconsolidation process and standard extinction are fundamentally different processes with different behavioral effects, memory encoding effects, brain region effects, and molecular cascade effects [18, 19, 20, 21, 34, 38].

A.17. Kitamura et al. (2020) [128]: "Boundary conditions of post-retrieval extinction: A direct comparison of low and high partial reinforcement"

The study by Kitamura et al. [128] had the same basic design as experiment 2 of Schiller et al. [6], using three different-color squares as CSs, two of which (yellow and blue, CS1+ and CS2+) were paired with an electric shock US, and the third square (grey, CS–), the control stimulus, was never paired with the US.

The main aim of this study was to contribute the first direct comparison of how fear memories formed with low and high partial reinforcement respond to post-reactivation extinction (PRE) and to standard extinction (SE). Therefore, on Day 1, acquisition for one group of subjects was at 40% reinforcement (each CS+ presented 10 times, 4 of those reinforced) and for another group was at 80% reinforcement (each CS+ presented 10 times, 8 of them reinforced). For both groups, the CS– was presented 10 times without reinforcement. Each CS square was presented for 6 s with an inter-trial interval of 11±1 s. Fear was measured as SCR response.

On Day 2, all subjects initially experienced one unreinforced presentation of CS1+, reactivating the threat memory of that CS+ but not the threat memory of CS2+. That was followed by the appearance onscreen of a Likert scale for rating the subjective expectancy of the US during the previous unreinforced presentation of CS1+, with the onscreen question, "How much did you expect to receive a shock after the square you just viewed?" The authors also explain (p. 4), "Participants were instructed to give a verbal response to the experimenter when prompted so that it did not require movement." Next came a viewing of a 10-min video of nature scenery, and then all subjects immediately underwent extinction training, with all three CSs presented a total of 11 times unreinforced on Day 2, including the earlier reactivation presentation of CS1+. Thus for all subjects, on Day 2 the fear memory of CS1+ was subjected to PRE and the fear memory of CS2+ was subjected to SE.

Testing on Day 3 began with measuring spontaneous recovery through a re-extinction process, the unreinforced presentation of 6 of each CS+ and 7 CS-s. Then, after a 30-s pause, 4 unsignalled shocks were administered for reinstatement, and then an animal nature video was viewed for 10 min before recovery of fear was elicited by unreinforced presentation of 4 of each CS+ and 5 CS-s.

The central result reported by Kitamura et al. [128] is that for both the 40% and 80% threat memories, after PRE on Day 2 there was no Day 3 return of fear of CS1+ due to either spontaneous recovery or reinstatement. That is, SCR levels in response to CS1+ and CS– were statistically indistinguishable in Day 3 tests, whereas at the end of acquisition on Day 1, CS1+ was generating SCR levels about 5 times the level of CS–. Annulment of the threat memory of CS1+ by PRE at both 40% and 80% reinforcement is indicated by those data.

In contrast, for subjects who underwent SE on Day 2, SCR levels for CS2+ were significantly higher than for CS– in both spontaneous recovery and reinstatement on Day 3, showing that the threat memory of CS2+ had remained responsive and functional after SE.

Kitamura et al. [128] acknowledge that "Prediction error (i.e., mismatch between what is expected and what actually occurs) is thought to play an important role in the modification of memory" (p. 9). They provide discussion of PE in terms of latent causes to interpret their SE results (return of fear) but not their PRE results (annulment of fear). The latter can be accounted for as follows by the PE and latent cause framework of Section 3.3.1: On Day 2, the transition from the CS1+ reactivation cue to the video consisted of the US expectancy rating procedure described above. That began with the novel, unexpected onscreen appearance of a Likert scale, and the subject immediately understood that this novel onscreen material was an important part of the study. That understanding had two important effects: It preserved the latent cause that had been active to that point, and it primed the subject to regard anything novel that subsequently appears onscreen as an integral part of the study. Therefore, neither the US expectancy rating procedure nor the video that appeared next induced a change in latent cause. Therefore, each of those unexpected perceptions created a PE experience, destabilizing the target memory and allowing its unlearning and annulment, according to the framework proposed in Section 3.3.1. (In clear contrast, the post-reactivation transitions carried out by Zuccolo and Hunziker [28] and by Houtekamer et al. [129], reviewed in Sections B.11 and B.13, respectively, caused subjects to view the coming video as not being part of the study, which induced a change of latent cause of the video and, therefore, neither a PE experience nor destabilization nor memory updating occurred, making these studies replication failures.)

Kitamura et al. [128] (p. 13) state, "Given that both low and high partial reinforcement groups did not demonstrate recovery of conditioned responses to the PRE stimulus, we conclude that a stronger memory formed as a result of a high partial reinforcement schedule does not function as a boundary condition of post-retrieval extinction." However, as noted in Section 5, the assumption that higher partial reinforcement forms a stronger memory is questionable [28, 131]. Comparing their SCR data with their data on US expectancy during reactivation, Kitamura et al. show that US expectancies for the reactivation CS1 after 40% and 80% acquisition reinforcement were equal, but the SCR level in response to the reactivation CS1 was higher for 80% reinforcement, indicating differential responses of declarative knowledge and affective arousal in relation to CS1 threat. Certainly Kitamura et al. have shown that the post-retrieval extinction process is effective across a wide range of partial reinforcements, a conclusion that is expanded further by noting also that even threat memories formed by 100% reinforcement were nullified via PRE with humans in the studies by Agren et al. [34], Thompson and Lipp [74], and Chen et al. [110], reviewed in Sections A.3, A.11, and A.18, respectively. Concrete procedures varied considerably across all of those studies (e.g., Agren et al. had lamps lighting within a scene; Thompson and Lipp had fear-relevant stimuli and US-reactivation; Chen et al. paired each CS+ with two shocks of variable timing), but the procedures in each study, including that of Kitamura et al. [128], created the same three experiences: target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.18. Chen et al. (2021a) [110]: "Destabilizing different strengths of fear memories requires different degrees of prediction error during retrieval"

Chen et al. [110] aimed to find "whether different strengths of conditioned fear memories require different degrees of PE during memory reactivation in order for the memories to become destabilized" (p. 1). In the first of two experiments, Day 1 acquisition training consisted of 6 presentations of the CS+ (an image of a colored geometric object), each accompanied by two US electric shocks (each with 0.2 s duration), for 100% reinforcement. No shock accompanied the 6 presentations of CS– (an image of a different geometric object with different color). The inter-trial interval was 16±1 s. Fear was measured as both fear-potentiated startle (FPS) and skin conductance response (SCR).

For one group of subjects, the two shocks always had the same timing within each 6-s CS+ (coming at 4.8 s and 5.8 s). The authors define that acquisition training as creating a CS+ memory having an "ordinary" level of fear for this "predictable-shock" group. For a second "unpredictable-shock" group of subjects, each 6-s presentation CS+ presentation had a different timing of the two shocks, creating an "enhanced" level of fear due to the uncertainty of when shocks would occur.

On Day 2, for each group the CS+ was presented once for 6 s to reactivate the target fear memory, followed by seeing onscreen a 10-min video segment from a BBC nature documentary, and then an extinction process ensued. For some subjects in the predictable-shock group, the single CS+ presentation was accompanied by two shocks having the same timing as in Day 1 acquisition, so no PE was created by this CS+ presentation to the predictable-shock/no-PE group. For other subjects in the predictable-shock group, the single CS+ presentation was accompanied by only one shock, mismatching the reactivated memory's expectation of two shocks and creating a PE experience along with reactivation for this predictable-shock/PE group. For subjects in the unpredictable-shock group, the single CS+ presentation was accompanied by only one shock, creating a milder mismatch for this unpredictable-shock/PE group than for predictable-shock/PE group subjects who received one shock on Day 2.

Memory testing on Day 3 consisted of spontaneous recovery and reinstatement tests. Of the three Day 2 reactivation conditions described just above, the CS+ threat memory was annulled only for subjects in the predictable-shock/PE group. Fear returned for the predictable-shock/no-PE group and the unpredictable-shock/PE group. In interpreting these results, the authors regard the different degrees of memory mismatch in the three conditions as different degrees of PE. They infer that PE consisting of one CS+ accompanied by one shock was strong enough to destabilize the "ordinary" level of fear in the predictable-shock CS+ memory, but not strong enough to destabilize the "enhanced" fear memory in the unpredictable-shock CS+ memory. In the PE analysis framework defined in Section 3, the different degrees of memory mismatch are not different degrees of PE. Rather, a mismatch condition that did not cause destabilization and annulment did not generate a PE experience, as depicted in Figure 1A.

A potentially instructive exercise in PE analysis consists of examining the predictable-shock/no-PE condition of Chen et al. [110] in relation to the PE analysis given above in Sections A.1 and A.2 for the retrieval-extinction results of both Schiller et al. [6] (experiment 1) and Oyarzún et al. [36]. According to the latter PE analysis, no PE was created by the single CS+ reminder being unreinforced because the target fear memory, which had been formed with partial (38%) reinforcement, did not expect the US to accompany the CS+ on every presentation. After that no-PE reminder, PE was then created by seeing the TV video appear next rather than the expected series of CS presentations and inter-trial intervals, and this PE caused memory destabilization, allowing updating and annulment by the extinction training that followed. In contrast, Chen et al. [110] found that their no-PE reminder, which also was followed by a 10-min video, did not result in updating or annulment, rather fear returned strongly in both spontaneous recovery and reinstatement. An explanation is needed for those opposite results from no-PE reminder conditions.

The explanation proposed here is an application of the latent cause analysis framework delineated in Section 3.3.1, focused on the different effects of 38% versus 100% reinforcement during acquisition, producing threat memories with different degrees of predictive uncertainty of US occurrence, and therefore different responses to the same reactivation conditions. The 100% acquisition pattern used by Chen et al. [110] produced predictive certainty. Then the reinforced CS+ reactivation on Day 2 matched (reproduced) the 100% acquisition pattern with zero ambiguity, so the reactivation CS+ was perceived as definitely having the same latent cause as the target learning, which decisively established that latent cause as the active latent cause. The video appeared next in that context, and due to the definiteness of the current latent cause and its lack of any recognizable relationship to the completely novel video, the video registered as evidence of a new, different latent cause, not an error of prediction in relation to the latent cause of CS+ (the latent cause in the target learning). Therefore, no destabilization occurred. In contrast, for the 38% acquisition reinforcement used by Schiller et al. [6] and Oyarzún et al. [36], predictive uncertainty of US occurrence was strong. The single unreinforced CS+ reactivation did not create PE, but, because it was ambiguous as to whether it had the same or different latent cause as the acquisition training, it did create further uncertainty about the already uncertain active latent cause. In that context of latent cause uncertainty and ambiguity, the appearance of the video registered subjectively as possibly relevant to the latent cause in the target learning, preserving the relevance of that latent cause, so the video was experienced as PE relative to the target learning's latent cause, destabilizing the target memory and allowing its subsequent updating and annulment. Thus, the PE evaluation framework of Section 3 is consistent with the opposite results and provides a cogent candidate explanation of them.

Experiment 2 of Chen et al. [110] consisted of two other groups undergoing the same unpredictable-shock acquisition training, which presumably forms a stronger fear memory than the predictable-shock training, and which in experiment 1 did not result in annulment with a one-shock mismatch reminder on Day 2. Experiment 2 tested two ways of attempting to increase the amount of Day 2 reactivation mismatch for that acquisition training. One way was a reactivation by CS+ with no accompanying shock, which again did not produce annulment, i.e., fear returned in response to reinstatement about as strongly as it did for the one-shock reactivation in experiment 1. Evidently, the changing timing of the two shocks accompanying each CS presentation throughout the unpredictable-shock acquisition training created a level of predictive uncertainty relative to which even the no-shock reactivation was still not a large enough mismatch to register as PE (depicted in the left-hand section of Figure 1A), as the one-shock reactivation had also failed to do.

The other attempted increase of reactivation mismatch in experiment 2 was a doubling of the reactivation procedure, i.e., two successive presentations of the CS+ accompanied by one shock, instead of one presentation as in experiment 1. This did result in annulment, i.e., no return of fear at Day 3 testing. Doubling the one-shock reactivation evidently was a large enough mismatch to register as PE and destabilize the target memory (depicted in the middle section of Figure 1A).

The authors conclude (p. 15), "our results indicate that PE is critical and necessary to trigger destabilization in retrieval-extinction". This study also serves well to illustrate the use of the conceptualization introduced above in Section 3.2.3, distinguishing between memory mismatch at reactivation as a procedural, objective condition and prediction error (PE) as a subjective condition that may or may not be produced by a particular mismatch condition. Chen et al. (2021a) describe their various experimental conditions as having different amounts of PE (e.g., on p. 15: "we tested two approaches to increasing the degree of PE in experiment 2"), which has been the customary view among researchers who consider PE. However, according to the conceptualization proposed here, those are not different degrees of PE, rather they are different degrees of reactivation mismatch, and only the mismatches that resulted in memory updating and annulment produced PE, inducing destabilization, as depicted in Figure 1A.

All things considered, this study adds its assent to the proposition that the annulment of an emotional learning through memory reconsolidation occurs as a result of three experiences: target learning reactivation, a concurrent PE experience, and a counter-learning experience.

A.19. Chen et al. (2021c) [133]: "Gender difference in retrieval-extinction of conditioned fear memory"

The first detection of gender differences in response to the retrieval-extinction procedure was reported by Chen et al. [133], who used the template of experiment 1 by Schiller et al. [6]. Day 1 acquisition training consisted of 6 presentations of the CS+ (a stereoscopic, three-dimensional image of a colored geometric object), 3 of which were accompanied by US electric shock (0.2 s), for 50% reinforcement. No shock accompanied the 6 presentations of CS– (an image of a different geometric object with different color). Each CS was visible for 5 s and the inter-trial interval was 9±1 s. Fear was measured as skin conductance response (SCR).

On Day 2, subjects in the retrieval-extinction group viewed the unreinforced CS+ presented once for 5 s to reactivate the target fear memory, followed by seeing onscreen a 10-min "neutral video," and then an extinction process ensued, with each CS presented 6 times unreinforced in random order. For subjects in the standard extinction group, the procedure on Day 2 consisted of only the extinction training. Thus, the PE analysis for the two groups is the same as delineated in Sections 3.3 and A.1 above for the Schiller et al. [6] study, according to the framework of PE analysis in Section 3.

Testing on Day 3 consisted of measuring spontaneous recovery via re-extinction, followed by a reinstatement test that began with four unsignalled shocks separated in time by 1 s, followed by each CS presented 4 times unreinforced in random order.

Data analysis divided each of the two groups into females and males, which showed the two genders having similar effects for spontaneous recovery but a striking difference for reinstatement:

- “there was no fear spontaneous recovery in the [male and female] participants who received the retrieval-extinction training, while there was fear spontaneous recovery in the participants who received the traditional extinction training.” (p. 1087)
- “male participants who received retrieval-extinction or traditional extinction training showed a fear reinstatement effect, while female participants showed no fear reinstatement effect.” (p. 1087) After the reinstatement shocks, females in the retrieval-extinction group showed zero return of fear, whereas males in the retrieval-extinction group had the same significant SCR level in response to CS+ as males in the standard extinction group, i.e., the retrieval-extinction procedure did not prevent return of males’ fear any more than standard extinction did in response to reinstatement. Females in the standard extinction group also had significant return of fear after reinstatement.
- Data analyzed overall, with genders combined, show less return of fear in the retrieval-extinction group than in the standard extinction group in both the spontaneous recovery and reinstatement tests. However, the authors emphasize that the observed gender differences (and other possible individual differences of phenotype) possibly could explain much of the inconsistent results of studies of the retrieval-extinction procedure.

The gender differences of response to the retrieval-extinction procedure found by Chen et al. [133] are a significant development in MR research, differing from the absence of gender differences previously reported in retrieval-extinction studies [134, 136]. The PE specificity analysis framework in Section 3.3.2 provides a candidate explanation for the observed gender difference by recognizing that the acquisition memory has several components encoded in different memory systems, and that a CS+ reminder (in Day 2 reactivation and in Day 3 spontaneous recovery test) and a US reminder (in Day 3 reinstatement test) reactivate different subsets of those components. As viewed within that phenomenology, the Day 3 test data of Chen et al. [133] from the retrieval-extinction group show that the threat memory could no longer be reactivated by CS+ presentation for both males and females, but it still could be reactivated by a US reminder for males but not females. It is well established that US presentation reactivates all linkages to US threat memory whereas CS+ presentation reactivates only the US linkage of that particular CS+ [69, 79, 95, 134, 135]. Evidently, males formed more US linkages than only the CS+ in the Day 1 acquisition training, whereas females formed only the CS+ linkage, with the result that Day 2 reactivation by CS+ destabilized all US linkage for females but not for males, and therefore Day 3 presentation of the US still reactivated the threat memory for males but not for females. This inferred difference in the memories produced by the acquisition training could be due to effects of cultural conditioning and/or genetics.

Thus, the PE evaluation framework of Section 3 applied to this study by Chen et al. [133] generates the prediction that US threat memory is immediately generalized during Day 1 acquisition by males but not females, with the result that threat memory reactivation by CS+ on Day 2 does not destabilize all of the US linkages in males, but does destabilize all of the US linkages in females. This study by Chen et al. is itself empirical support for this explanation, but this explanation could also be tested by replicating the study with the one difference of presenting the US instead of the CS+ for Day 2 reactivation, which should destabilize and allow annulment of all US linkages in both males and females, eliminating the return of fear after reinstatement also in males.

This study by Chen et al. [133] again shows that when the behavioral updating does result in full annulment of an emotional memory, the procedure that had that effect created a series of three experiences: target learning reactivation, a concurrent PE, and a counter-learning experience.

A.20. Chen et al. (2021b) [127]: “Retrieval-extinction as a reconsolidation-based treatment for emotional disorders: Evidence from an extinction retention test shortly after intervention”

In experiment 1 of Chen et al. [127], the protocol of experiment 2 of Schiller et al. [6] was used to measure, for the first time, the responsiveness of the fear memory soon, on the same day, 3 h after completion of the Day 2 retrieval-extinction procedure, by two unreinforced presentations each of CSa+, CSb+, and CS–. That exploration was extended in their experiment 2, which added another test at 12 h, with and without sleep during that 12-h period. In both experiments, CSa+ was the unreinforced reminder presented once 10 min before the extinction procedure. In experiment 2, another addition was an unreinforced single presentation of CSb+ 10 min after the extinction procedure, forming a concurrent extinction-retrieval (reversed) protocol for the CSb+ fear memory. The fear response tests in experiment 2 at 3 h and 12 h evaluated the short-term effects of each of those two protocols, retrieval-extinction and extinction-retrieval. In each of the three short-term test conditions (3 h, 12 h, and 12 h after sleep), there was no non-reminded fear memory, i.e., the CSa+ reminder was always presented 10 min before extinction and the CSb+ reminder was always presented 10 min after extinction. Fear was measured as skin conductance response (SCR).

For both experiments, Day 1 acquisition training consisted of 6 presentations of CSa+, CSb+, and CS- for 5 s each and an inter-trial interval of 11±1 s, with 4 of the CSa+ and CS+b presentations paired with an electric shock US for 67% reinforcement.

On Day 2, the 10-min period before (and after, in experiment 2) the extinction procedure is defined in this way by the authors (p. 6): "A resting period of 10 min was inserted between reactivation and extinction. During the break, participants watched an excerpt from [the BBC documentary] Planet Earth….During all sessions of the experiment, the shock stimulator was set to the "on" position and the SCR was recorded continuously." Subjects remaining connected to electronic equipment set to "on" during the 10-min delay periods presumably led them to regard the 10-min period as being part of the study. That would maintain the active latent cause and thereby allow a PE experience to be created by the adjoining mismatches of the expected series of CSs not appearing and instead, the unexpected video appearing, destabilizing the CS+ threat memory, as explained in Section 3.3.1.

Testing of fear memory response on Day 3 consisted of re-extinction (6 each of unreinforced CSa+, CSb+, and CS–) to measure spontaneous recovery, followed after 1 min by 4 unsignalled US shocks for a reinstatement test (4 each of unreinforced CSa+, CSb+, and CS–).

In experiment 1, the CSa+ reminder evoked no return of fear in either test on Day 3, indicating that the CSa+ fear memory had been updated and nullified by retrieval-extinction. This is therefore a successful replication of the study by Schiller et al. [6]. The non-reminded CSb+ fear response returned significantly after reinstatement, though it did not show spontaneous recovery in response to CSb+ presentation, indicating that the CSb+ extinction memory had remained dominant on Day 3 until the unsignalled US presentations reinstated the CSb+ fear memory. A relatively weak acquisition training with only four CS+/US pairings is presumably why the standard extinction of CSb+ was strong enough to persist in that way. Oyarzún et al. [36] had six CS+/US pairings and Li et al. [113] had twelve.

In experiment 2, there was no return of CSa+ fear or CSb+ fear in any of the Day 3 tests, i.e., both the retrieval-extinction protocol and the extinction-retrieval protocol prevented the return of fear, which is a successful replication with human subjects of earlier animal studies of the extinction-retrieval protocol [75, 76, 77]. It is valid, therefore, to interpret the results of experiment 2 of Chen et al. [127] as indicating that neither of the protocols produced an extinction memory competing with an acquired fear memory, that therefore neither protocol functioned as a standard extinction training, and that each of them functioned by recruiting the MR process and subjecting both the CSa+ fear memory and the CSb+ fear memory to the MR process of destabilization, updating, and nullification.

However, as noted in Section 4.2, viewed as a series of procedural steps, the extinction-retrieval protocol has been regarded by memory researchers as conceptually incompatible with recruitment of MR [52, 69, 73]. That opinion appears to be shared strongly by Chen et al. [127], who interpret their short-term, 3 h and 12 h memory test data as indicating that the extinction-retrieval protocol caused standard extinction, not MR. (The CSa+ memory showed return of fear at 3 h and 12 h, but not at 12 h with sleep, whereas the CSb+ memory never showed return of fear at 3 h, 12 h, or 12 h with sleep.) That is the necessary interpretation if it assumed a priori that an extinction-retrieval procedure can cause only extinction, not updating via MR. However, the conceptual feasibility of the extinction-retrieval protocol recruiting MR and updating a fear memory is shown in Section 4.2. That account also addresses how the proposed extinction-retrieval phenomenology can explain the short-term test results of Chen et al. [127] (including the important finding that sleep may be necessary under certain conditions for behavioral updating to have its full, final effect). The conjectured phenomenology creates the three experiences of target learning reactivation, a concurrent PE experience, and a counter-learning experience, maintaining the universality of that set of experiences as the primary requirement for emotional memory annulment to occur. The above PE evaluation of the retrieval-extinction protocol of Chen et al. [127] implies that it too created those three experiences.

Appendix B. 14 non-replication studies of human behavioral updating

The following reviews provide details of each non-replication study's procedure and results that are relevant to identifying the experiences induced in subjects, including prediction error (PE) experiences. For full details, the original journal article must be consulted. Examples of items not included here are baseline measurement procedures, habituation procedures, statistical analyses, subject demographics, and subject inclusion and exclusion criteria.

B.1. Soeter & Kindt (2011) [33]: "Disrupting reconsolidation: Pharmacological and behavioral manipulations"

Experiment 2 of Soeter & Kindt [33] utilized the protocol of experiment 2 of Schiller et al. [6] (reviewed in Section A.1) with two main variations: The two CS+ images paired with electric shock US were fear-relevant stimuli (gun and spider) rather than neutral geometric shapes (but the CS– image was neutral, a mug); and rather than 38% reinforcement during Day 1 acquisition training, four of the five training trials of CS1+ and CS2+ were reinforced, which could be defined as 80% reinforcement, but the four reinforced trials were chunked together after an initial unreinforced trial, resulting presumably in very nearly the same levels of contingency certainty and US expectancy in response to a subsequent CS+ presentation as a 100% reinforcement. Subjects "were told to learn to predict whether an electric stimulus would occur or not. In addition, participants were required to rate their distress (i.e., anxiety, tension, or nervousness) during the presentation of each slide by shifting a cursor on a visual analog scale and push the left mouse button within 5 sec following stimulus onset…" (p. 365). Also, fear-potentiated startle (FPS) response was measured in addition to SCR, with the startle probe presented 1 s before CS offset during acquisition training.

On Day 2, one CS1+ image was presented for 8 s, unreinforced, to reactivate the threat memory of only that image. Subjects rated their distress as during acquisition. That unreinforced presentation was not a mismatch of the acquisition conditions, so it did not create a PE experience and did not destabilize the reactivated fear memory. Next, as explained by the authors (p. 365), "Subsequent to memory reactivation, a 10-min break was inserted (Schiller et al. 2010). During the break, participants were offered magazines to read. Extinction learning immediately followed."

Testing for return of fear on Day 3, was done first by re-extinction to detect spontaneous recovery, followed by a reinstatement test beginning with 3 unsignalled USs, followed by a reacquisition test, followed lastly by a test of fear generalization. For each of those four tests, responses were measured by FPS, SCR, and subjective distress rating (with the exception that distress data is not presented for re-extinction). Significant return of fear is indicated by all but two of those eleven tests results: No differential return of fear was measure by FPS in re-extinction (spontaneous recovery) and by SCR in reinstatement. Given the preponderance of data indicating that the threat memory was not nullified and reconsolidation was not induced, the authors maintain that strong extinction on Day 2 explains the absence of spontaneous recovery in FPS data, and they note that though the SCR data shows no differential return of fear after reinstatement, the absolute level of SCR is significantly higher than at the end of extinction on Day 2, indicating a significant affective response (though not necessarily a fear response, as SCR occurs bivalently). The authors conclude regarding the observed return of fear (p. 363), "Overall, the present findings stand in sharp contrast to those reported by Schiller et al. (2010)."

For identifying the cause(s) of this study's non-replication, the best approach may be a comparison with the subsequent study by Golkar et al. [188] (reviewed in Section A.10), which used an identical procedure for Day 1 acquisition training and Day 2 reactivation and 10-min delay (with one important exception identified below), but which succeeded in producing updating and annulment of the threat memory. The cause of non-replication must be one or both of the two specific ways in which the procedure used by Soeter and Kindt [33] differed from that of Golkar et al.

First, continually onscreen during all phases of the study by Soeter and Kindt [33] was a horizontal analog scale that each subject used during the first 5 s of each CS presentation for indicating level of subjective distress from 0 to 10 by sliding the cursor and clicking a mouse button. Golkar et al. did not have subjects make ratings of their subjective experience. Second, for post-reactivation extinction training on Day 2, Soeter and Kindt used the same standard extinction procedure that was used in the previous successful retrieval-extinction study by Schiller et al. [6], whereas Golkar et al. used a novel, vicarious extinction procedure in which the subject views a 24-min video of a person viewing a computer screen with a relaxed demeanor while seeing the same CSs, unreinforced, appearing.

The difference in extinction procedure cannot be the cause of non-replication because Soeter and Kindt [33] used the extinction procedure that was previously used successfully by Schiller et al. [6]. Therefore the strongly probable cause of non-replication is subjects participating under the continual requirement of mindful attention toward subjective distress and declarative knowledge of the CS+/US contingency. Soeter and Kindt addressed that possibility, as did several other study authors. For example, Oyarzún et al. [36] explained (p. 4), "This continuous evaluation of the association between CS-US could have overstrengthened the more conscious association between CS-US. This cortical representation of CS-US might have elicited fear responses in the amygdala even if the association would have been effectively disrupted by the behavioral manipulation." Steinfurth et al. [100] explained further (p. 341) that "emphasizing explicit knowledge of the CS–UCS relationship has been shown to alter the nature of fear learning…and the neural substrates mediating this learning (Coppens et al., 2009; Funayama et al., 2001)." Chen et al. [110] explained their experimental design considerations (p. 15), "we do not use declarative US-expectancy ratings to measure PE in the current study, because we found that the online US-expectancy ratings may affect the sensitivity of SCR to conditioned fear in a pilot study." The matter of possible distortions of findings caused by US expectancy ratings was important enough that Warren et al. [148] carried out a study specifically to compare results with and without such ratings, but neither condition replicated the updating and annulment effect, so a useful comparison did not emerge (reviewed in Section B.5). Kredlow et al. [53] pointed out (p. 46), "Our meta-analysis of PRE studies in healthy human samples found that studies using CS-UCS expectancy ratings during acquisition, potentially enhancing CS-UCS contingency awareness and hippocampal-dependence, had smaller effects than studies that did not use expectancy ratings (Kredlow et al., 2016)."

The probability that continual US expectancy distress ratings were the cause of non-replication by Soeter and Kindt [33] is enhanced further by the fact that none of the 20 successful behavioral updating studies listed in Table 1 used continual US expectancy ratings. Seventeen of those studies did not involve US expectancy ratings at all, and in the other three, US expectancy ratings were made only in hindsight after, never during, an experimental procedure: Thompson and Lipp [74] had subjects rate the affective valence of CSs and US, but ratings were made only after each experimental phase was completed, rather than continuously during each phase, which significantly reduced or eliminated the interfering effect of the subjective ratings. Kitamura et al. [128] had subjects rate US expectancy only once, after reactivation CS offset on Day 2. Similarly, Chen et al. [127] had subjects rate US expectancy only once, at the end of Day 2 after a short-term test of fear responding.

It is noteworthy that according to the analysis above, the study by Soeter and Kindt [33] was a replication failure not due to a stronger fear memory created by using fear-relevant stimuli and a much higher reinforcement percentage than in the study by Schiller et al. [6], which are some of the possible explanations given by the authors. Indeed, in subsequent years, behavioral updating and annulment of fear memory were demonstrated using high reinforcement percentage in four studies [34, 110, 125, 128], and with both fear-relevant stimuli and high reinforcement percentage in two studies [74, 188].

B.2. Golkar et al. (2012) [131]: "Are fear memories erasable? Reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli"

Golkar et al. [131] reported two experiments, both patterned on experiment 2 of Schiller et al. [6], and both failed to replicate the annulment of a threat memory by the retrieval-extinction protocol.

For present purposes, first considering experiment 2 of Golkar et al. [131] will be clearest. It differed slightly from experiment 2 of Schiller et al. [6] in these parameters: Acquisition training used partial reinforcement of 50% (instead of 38%), each CS presentation had duration of 6 s (instead of 4 s), and the inter-trial interval was 14±4 s (instead of 11±1 s).

Due to partial reinforcement, the single unreinforced CS+ presentation on Day 2 did not create a PE experience. The content of the "video clip" viewed by subjects during the post-reactivation "10-min break" is not described by Golkar et al. [131], nor is the manner of transitioning to the video clip, so those factors cannot reliably be evaluated for how they may have contributed to a change in latent cause that would have prevented PE and destabilization. If in the transition subjects were led to view the 10-min period as a "break," i.e., a period of rest and relaxation outside of the study, then that would have activated a new latent cause and prevented a PE experience from occurring. The result would have been no destabilization, updating, or annulment of the fear response, as observed. This is the most likely explanation of this non-replication, as understood in terms of the PE and latent cause framework of Section 3.3.1.

Experiment 1 of Golkar et al. [131] differed from their experiment 2 by using fear-relevant stimuli (3 different images of a man's threatening face) and fear-potentiated startle response to measure fear in addition to SCR. This study also did not find annulment of the fear memory. The use of fear-relevant stimuli in this experiment almost certainly had the effect of making even clearer the irrelevance of the video to the latent cause in the acquisition training than in experiment 2, thereby making the video even more likely to induce a change of latent cause than in experiment 2.

Thus, for the study by Golkar et al. [131] it is possible to account for the non-replication of results in terms of how PE experiences are created and latent causes are changed, as defined in Section 3, but the lack of key information makes this a more speculative analysis than for most of the other non-replication reviews here in Appendix B.

B.3. Kindt & Soeter (2013) [179]: "Disrupting reconsolidation: Pharmacological and behavioral manipulations"

Kindt & Soeter [179] used the protocol of experiment 1 of Schiller et al. [6] with two main variations: the CS+ and CS– images were fear-relevant stimuli (spiders) rather than neutral geometric shapes, and the CS+ was paired with electric shock US during Day 1 acquisition training on 75% of CS+ presentations (6 of the 8 trials, each shown for 8 s with an inter-trial interval of 20±5 s) rather than 38%. Also, in addition to SCR, fear-potentiated startle (FPS) response was measured, and subjects made subjective ratings of US expectancy in response to every CS presentation through the study, in same manner as described above in the review of the study by Soeter and Kindt [33].

On Day 2, for subjects in the reactivation-extinction condition, the CS+ image was presented for 8 s, unreinforced, to reactivate the threat memory of that image. Next, as explained by the authors (p. 45), "Subsequent to memory reactivation, participants were detached from the experimental set-up and a resting period of 10-min was inserted (Schiller et al., 2010). During this break, participants were offered magazines to read. Prior to extinction learning, participants were again attached to the experimental apparatus…." After 10 min, extinction training began, consisting of 11CS+, 12CS–, and 11 noise-only trials, all unreinforced and randomized. Subjects in the extinction-only condition began Day 2 with the 10-min period with magazines, followed by extinction, now with 12CS+.

Testing for return of fear was done on Day 3, first by re-extinction to detect spontaneous recovery, followed by a reinstatement test beginning with 3 unsignalled USs. Both FPS and SCR, as well as US expectancy ratings, showed significant return of fear of the reminded CS+ in both spontaneous recovery and reinstatement, and the authors conclude that (p. 43) "extinction learning within the reconsolidation window did not prevent the recovery of fear on multiple indices of conditioned responding (startle response, skin conductance response and US-expectancy)."

The highly probable causes of non-replication seem particularly clear for this study. First, this study had subjects engaged in the same continuous ratings of their subjective US expectancy as in the study of Soeter and Kindt [33], reviewed above in Section B.1, where that cause alone was deemed the probable cause of non-replication. Here in addition is clear information, quoted above, about how the researchers carried out the Day 2 transition from CS+ offset to the 10-min period. Detaching subjects from the electronic equipment for SCR, FPS, and US and offering magazines to read is certain to mean to them that this 10-min period is not part of the study, which is a change of context and a change of latent cause immediately after CS+ offset. As explained in Section 3.3.1, that prevents the creation of two PE experiences that would otherwise occur due to the CS+ not being followed by the expected series of CS presentations and due to the unexpected onscreen appearance of the video. The reactivation CS+ presentation being unreinforced also did not create a PE experience because that was not a mismatch of the acquisition conditions. Without a PE experience occurring, the target memory was then not in a destabilized condition when the extinction training began after the 10-min break. Therefore, no updating or annulment resulted, and presumably that would have been the result even if no US expectancy ratings were made.

Understood in that way, this study's replication failure was not due to either of the study's more obvious differences from the study by Schiller et al. [6], i.e., use of fear-relevant stimuli and the higher reinforcement percentage. As detailed in Section B.1, six studies have produced behavioral updating and annulment of threat memory using procedures with high reinforcement percentage, two of them also with fear-relevant stimuli.

B.4. Meir Drexler et al. (2014) [149]: "Effects of postretrieval-extinction learning on return of contextually controlled cued fear"

Aiming to test the post-retrieval extinction protocol for the first time on a contextually controlled cued fear memory acquired with fear-relevant conditioned stimuli in humans, and using SCR as an index of fear, Meir Drexler et al. [149] found the same degree of return of fear in both the standard extinction group and the reactivation+extinction group. The authors conclude (p. 479), "the present study failed to find a beneficial effect of postretrieval-extinction…."

They also speculate about three mnemonic effects that might possibly have contributed to causing the replication failure:

- "Fear-relevant stimuli may have led to a stronger fear memory, which was not as vulnerable as the neutral shapes [used in other studies]." (p. 479);
- reactivation by CS+ presentation on Day 2 had a different visual context from that of the Day 1 acquisition training;
- the extinction training on Day 2 was done in a visual context different from that of the Day 1 acquisition training.

No consideration is given, however, to the empirical finding that destabilization requires a PE experience, as reviewed in Section 3.1. Carrying out the PE analysis as defined in Section 3 arrives at an explanation for the non-replication that is somewhat stronger than speculation:

Acquisition training consisted of 16 presentations of each of three CS images for 8 s, with 75% reinforcement of each of the two CS+ images by an electric shock US. The inter-trial interval was 10±2 s.

For the reactivation+extinction group, Day 2 reactivation consisted of an unreinforced, 30-s presentation of a CS+. The absence of the US did not create a PE experience because of partial reinforcement in acquisition. The 30-s duration is the parameter most important to evaluate. It was one of the early findings of MR research that a reactivation presentation that is far longer than the acquisition presentations does not destabilize the target memory, rather it functions as an extinction learning of a new, separate memory of no-threat [4, 82, 191]. Even more directly relevant is the study by Hu et al. [150] reviewed above in Section A.13, which systematically varied the duration of the Day 2 reactivation presentation and showed that destabilization resulted from a 1-s and 4-s presentation, but not from a 30-s and 3-min presentation. Evidently a change of perceived latent cause occurs when the reactivation duration is sufficiently longer than any presentation duration in acquisition.

It is strongly probable, therefore, that the 30-s reactivation used by Meir Drexler et al. [149] did not create a PE experience and therefore did not destabilize the target memory, rather it began the formation of an extinction learning. The effects of any next steps of procedure would now be on that extinction memory trace, according to the principle of trace dominance [183] that emerged from MR research to account for consistent observations that the next step of procedure acts upon the currently active memory trace, whether it be a destabilized memory undergoing reconsolidation or an extinction memory. In other words, the extinction training that ensued after the 10-min break further developed the extinction learning already underway, rather than updating the target threat memory.

Thus it is with some degree of reliability that the cause of this study's replication failure can be identified as a reactivation procedure that induced extinction learning rather than destabilization. This conclusion could be tested by repeating the study using an 8-s instead of 30-s reactivation duration, which should result in annulment of the target memory in the reactivation+extinction group, provided the transition from Day 2 CS+ offset to 10-min delay period is done in such a way that subjects regard the 10-min period as fully part of the study.

B.5. Warren et al. (2014) [148]: "Human fear extinction and return of fear using reconsolidation update mechanisms: The contribution of on-line expectancy ratings"

As in experiment 2 by Schiller et al. [6], Warren et al. [148] had two CS+s that were paired with an aversive US and a third CS– that was not paired. The CS+a, CS+b and CS– were geometric shapes presented onscreen for 6 s. Day 1 acquisition training differed significantly from the acquisition parameters of Schiller et al., however: Each of the CSs was presented 12 times and the aversive US was a 250 ms airblast directed at the larynx and delivered on all CS+ trials, i.e., 100% reinforcement. The inter-trial interval was randomized between 9 and 22 s. Fear was measured as acoustic fear-potentiated startle (FPS) response (for which subjects wore headphones).

This study had a 2X2 design: with and without a memory retrieval cue (unreinforced CS+) 10 min before extinction on Day 2, and with and without subjects' US expectancy ratings made using a three-button keypad on every CS presentation throughout the study in the first 3 s after CS onset. Revealing the effects of continual US expectancy ratings was one of the study's main goals due to the possibility that replication failures by Soeter and Kindt [33] and Kindt and Soeter [179] were due to the continual US expectancy ratings in those studies.

On Day 2, the CS+a was presented once, unreinforced, to subjects in the Retrieval group, but not to subjects in the No Retrieval group. That unreinforced CS+a was definitely a mismatch of the 100% acquisition training and would reasonably be assumed to create a PE experience that would destabilize the target memory. Then (p. 168), "Ten minutes after the Retrieval trial, participants were administered the Extinction Training session." There is no mention of video or magazines, which presumably indicates that subjects simply waited during the 10-min period and were informed that they would wait either by words onscreen, by a voice heard in the headphone, or by a person entering the room.

On Day 3, the return of fear was tested by re-extinction to detect spontaneous recovery, followed by a reinstatement test initiated by four unsignalled applications of the airblast US. The FPS data show a failure to achieve updating and annulment of the target fear memory in the Retrieval groups. All four groups showed significant spontaneous recovery of fear in Day 3 re-extinction, implying that no destabilization or reconsolidation occurred in any of the groups.

The FPS data also show that use of the keypad to make US expectancy ratings increased amygdala-generated fear in all phases of the study, confirming previous concerns, reviewed in Section B.1 above. With the keypad, the return of fear was smaller in the Retrieval group as compared to the No Retrieval group. Without the ratings keypad, the opposite was found: return of fear was greater in the Retrieval group as compared to the No Retrieval group. The not-destabilized target memory undergoing standard extinction 10 min after reactivation resulted in a stronger response on Day 3 than when no reactivation had happened before extinction.

In discussing possible causes of the absence of a reconsolidation updating effect, the authors state (p. 171), "Our current study used 100% reinforcement on the CS+ trials during fear acquisition….The studies by Schiller et al. (2010) and Oyarzun et al. (2012), however, employed a partial reinforcement schedule (37.5%). It is possible that the lack of a disruption of reconsolidation (as evidenced by spontaneous recovery observed in all groups of the present study) may be due to stronger conditioning and subsequent resistance to reconsolidation update."

In considering that explanation, it is necessary to account for the fact that three studies with 100% reinforcement in the acquisition training have achieved updating and annulment:

- Agren et al. [34]: 16 CS+ acquisition trials (reviewed in Section A.3)
- Thompson and Lipp [74]: 8 CS+ acquisition trials (reviewed in Section A.11)
- Chen et al. [110]: 6 CS+ acquisition trials (reviewed in Section A.18)

Therefore, understanding the replication failure with the no-keyboard groups of Warren et al. [148] is a matter of understanding why their target fear memory was not destabilized in the same manner as in the three studies listed above, i.e., by one unreinforced CS+ presentation. To that end, the study by Chen et al. [110] seems directly relevant (reviewed above in Section A.18). Chen et al. created two different fear memories of different strength, though each was acquired with 100% reinforcement. The milder memory was created with the same US timing on every acquisition trial, and it was destabilized by a single unreinforced CS+ presentation. The stronger memory was created with a different and therefore uncertain US timing on every acquisition trial, and it was not destabilized by a single unreinforced CS+ presentation, but was destabilized by a series of two identical unreinforced CS+ presentations. Thus, it seems quite probable that Warren et al. were correct about their target memory not destabilizing due to its strength. Their study, like that of Chen et al., seems to provide an example of a reactivation mismatch not being large enough to create a PE experience, as discussed in Section 3.3.1 and represented in the left portion of Figure 1A. That analysis could be tested by repeating the Retrieval/No Keypad group with two unreinforced CS+a presentations on Day 2, strengthening the mismatch sufficiently to produce a PE experience, destabilization, and nullification of the fear response.

As noted above, Agren et al. [34] created a fear memory using more acquisition trials than Warren et al. [148], yet only one unreinforced CS+ presentation sufficed to cause destabilization. Evidently, the fear memory created by Warren et al. was significantly stronger than that of Agren et al. even with fewer acquisition trials. That would imply, plausibly, that the physical impact of an airblast to the throat is experienced subcortically as a stronger and more fear-generating threat than the “unpleasant, but endurable” electric shock to the lower right arm used by Agren et al. With a sufficiently strong fear of the US, more presentations of an unreinforced CS+ are required in order to make it definite enough that the expected contingency is possibly no longer what it was, creating a PE experience.

B.6. Fricchione et al. (2016) [186]: “Delayed extinction fails to reduce skin conductance reactivity to fear-conditioned stimuli”

Noting that “Past successful experiments in reconsolidation blockade have produced a floor effect (i.e., the conditioned fear response is completely eliminated by the intervention)” (p. 2), these researchers aimed to create acquired fear responses significantly stronger and more resistant to updating than in previous studies in order to avoid the floor effect and thereby allow “the comparison of potential reconsolidation blocking techniques that differ in their efficacy” (p. 2).

Subjects with heightened fear responses were prepared by (a) pairing an electric shock US with vivid fear-relevant CSs (“8-s, high-definition video clips of three crawling tarantulas, each conspicuously different in appearance” (p. 2)), (b) testing for and selecting individuals with stronger fear responses to spiders (but not spider-phobic), and (c) including only subjects with strong differential SCR levels for two different spider videos during Day 1 acquisition training.

The Day 1 procedure consisted of these phases:

1. A viewing of a single still image of each of the three tarantulas;
2. a 5-min pause;
3. habituation: a viewing of one video clip of each of the three tarantulas in each of three different contexts (kitchen, bedroom, and office), with no US pairing;
4. acquisition: a viewing of 8 videos of each tarantula in the kitchen context, with 5 of the 8 paired with US shock (62.5% reinforcement) for two tarantulas, and no shock for one tarantula.

Each tarantula video clip began with 4 s of context alone (no tarantula), followed by 8 s of tarantula walking in that same context. The inter-trial interval was a black screen for 20±4 s. On reinforced trials, the US immediately followed the end of the video.

The Day 2 procedure began with a single unreinforced CS+ presentation (video of one of the shock-associated tarantulas), “and then waiting 10 min [no video] prior to presenting the remaining…extinction trials”. Extinction consisted of a series of unreinforced presentations of all three tarantulas, for a total of 10 each including the initial one before the 10-min wait. All tarantula video clips on Day 2 were in the bedroom context.

On Day 3, SCR measurements in both a renewal test and a reinstatement test showed that “the extinction delay had no measurable reconsolidation blocking effect on the fear-conditioned SCR….[A]n initial delay of 10 min between the first extinction trial and subsequent trials failed to diminish the targeted fear memory trace.”

In discussing possible causes of their replication failure, the authors give ample consideration to the PE requirement. They recognize that the single unreinforced CS+ presentation on Day 2 did not create PE due to the partial reinforcement in acquisition on Day 1, while noting that “Schiller et al. (2010) and Steinfurth et al. (2014) obtained positive results with similar rates of reinforcement as used in the present study. If, indeed, our conditioning procedure succeeded in establishing a more strongly conditioned response, it may be that a greater prediction error is needed to destabilize the fear memory.” (p. 7).

Applying the PE analysis framework from Section 3 suggests other possibilities:

- Fricchione et al. [186] had no video onscreen during their 10-min period of waiting. That 10-min wait was quite similar to the 5-min wait during the Day 1 procedure, and therefore was a sub-salient mismatch of acquisition conditions (see Figure 1A, left-hand section), so no PE experience and no destabilization was produced by this 10-min wait.
- Seeing only one CS+ during Day 2 reactivation, without being followed by another CS after 20±4 s, was also not a salient mismatch because subjects had been shown single presentations of each CS during habituation before acquisition training on Day 1. Without those single presentations on Day 1, the single Day 2 presentation

probably would have created a PE experience, destabilizing the target memory and resulting in successful updating, if the ensuing 10-min wait had begun without any communication from researchers defining it as a "break" occurring outside of the study. This is a testable prediction of the trial PE evaluation framework.

- The change of tarantula video context from kitchen during acquisition on Day 1 to bedroom on Day 2 may seem to constitute a mismatch that would create PE. However, the bedroom context also was viewed on Day 1 during habituation. In the habituation experience that preceded acquisition on Day 1, subjects learned that the context changes from one tarantula video clip to the next. Then, in the acquisition training, they learned that the context can also remain the same for many tarantula video clips. Those two qualitatively different patterns on Day 1 would probably be perceived as having different latent causes. Consequently, the single video clip at the start of Day 2 would have a strongly uncertain latent cause and therefore not register as a definite mismatch of a definite pattern learned on Day 1. Again, no PE experience would occur without a distinct mismatch.

Thus, according to the PE analysis framework from Section 3, the cause of the replication failure of Fricchione et al. [186] was not that PE was not strong enough for the strong fear of the target memory, but rather that no PE experience occurred at all. This view is in agreement with the authors' conjecture (p. 7) that "if a 100 percent reinforcement schedule had been used, then the unreinforced reactivation trial would have…caused the memory to become destabilized and amenable to reconsolidation blockade." In other words, regardless of the strength of fear response generated by the target memory, the memory will destabilize if reactivated with a suitable mismatch, one that is sufficiently salient and unambiguous to create a PE experience. The study by Chen et al. [110], reviewed in Section A.18, demonstrates exactly that effect by reactivating the same target memory with three different degrees of mismatch, and only the largest mismatch was sufficiently unambiguous relative to the acquisition training to produce destabilization.

B.7. Klucken et al. (2016) [190]: "No evidence for blocking the return of fear by disrupting reconsolidation prior to extinction learning"

The study by Klucken et al. [190] used the same retrieval-extinction protocol as in experiment 2 of Schiller et al. [6], but with some parameter changes: Acquisition training had partial reinforcement of 50% (instead of 38%) by the electric shock US, each CS presentation had duration of 8 s (instead of 4 s), and the inter-trial interval was 5.5–8 s (instead of 11±1 s). Also, fMRI BOLD images of brain activity were recorded in all phases, with subjects in an fMRI scanner, in addition to SCR measurements.

On Day 2, the reminder CS+ and the neutral CS– were presented for memory reactivation, followed by a "10-min break" and then an extinction training (10 reminder CS+, 10 CS–, 11 non-reminder CS+).

Day 3 test of re-extinction showed significant return of fear of the reminder CS+. This study failed to replicate annulment of fear memory by the retrieval-extinction protocol. There was no significant difference between subjects' responses to the reminded CS+ and the non-reminded CS+, as measured by both SCR and fMRI brain images of BOLD responses.

The authors speculate about numerous possible causes for this non-replication of the effect, but there is no consideration of the PE requirement or its possible role in the non-replication. The viewpoint of the present review is that as a rule, the PE requirement is the primary candidate for understanding and explaining non-replications of MR behavioral updating results. Applied to this study by Klucken et al. [190], analysis by the PE evaluation framework of Section 3 is as follows:

The observed sameness of SCR and fMRI data for the reminded and non-reminded CS+ mean that destabilization did not occur in response to the reminded CS+. That implies, in turn, that a PE experience was not generated. The unreinforced CS+ reminder presentation did not create a PE experience because it did not mismatch the acquisition training, as in the study by Schiller et al. [6]. But where the procedure of Schiller et al. did create PE, evidently the procedure of Klucken et al. [190] did not: According to the PE analysis in Sections 3.3 and A.1 for the study by Schiller et al., the post-reactivation PE experience on Day 2 that destabilized the fear memory was created by the appearance onscreen of a TV episode following the presentation of a single unreinforced CS+ reactivated the target memory. Klucken et al. also presented a video at that point, though it contained scenes of nature landscapes. No information is provided by Klucken et al. regarding the procedure for the transition from CS+ reminder offset to what they describe as a "10-min break" (p. 114) with the nature video. If that transition involved any communication to subjects that led subjects to regard the 10-min period as a "break," i.e., a period of rest and relaxation outside of the study, a change of latent cause is strongly probable. Likewise, disconnection of electronic equipment from the subject's body would send the same message and induce a change of latent cause, preventing creation of a PE experience that would otherwise occur and induce destabilization. Since no PE occurred, it has to be inferred that the elements of procedure that could have created PE were prevented from doing so.

This analysis, which is essentially the same as given for the study by Golkar et al. [131] in Section B.2, again is more speculative than for other non-replication studies in Appendix B, but it serves nevertheless to show that an explanation in terms of the PE evaluation framework of Section 3 is possible for this study's replication failure.

B.8. Kroes et al. (2017) [132]: "A reminder before extinction strengthens episodic memory via reconsolidation but fails to disrupt generalized threat responses"

Every CS+ and CS– presentation in the study by Kroes et al. [132] was a unique image of a bird or fish, never again presented. During acquisition training on Day 1, 20 images of members of each of those two categories were presented, and for the CS+ category, half of its 20 images were paired with an electric shock US to the wrist for 50% reinforcement. Each CS was visible for 4 s, with inter-trial interval of 12±1 s. In that way, a threat memory of one category (fish or bird) was created. The authors explain (p. 2), "As no single item repeats, only an item's membership of a specific category has predictive value and threat responses generalize to novel items of the reinforced category." Fear was measured as SCR.

Day 2 procedure began for subjects in the reminder group with presentation of a novel CS+ category image "that was the most prototypical exemplar of the CS+ category" (p. 3). Subjects "then watched a video (BBC Planet Earth) for 10 min before extinction training on novel unique items" (p. 10). Extinction consisted of 12 (reminder group) or 13 CS+ trials (no-reminder group) and 12 CS− trials, all without the US and all again unique images not seen previously.

Day 3 testing began with four unsignalled USs separated by 20, 30, and 25 s, for reinstatement of the threat response, followed after 15 s by randomized, unreinforced presentations of 10 CS+ and 10 CS– for measuring return of fear. Next were three different tests of episodic memory of the images seen on Days 1 and 2.

The same procedure with reminder on Day 2 was carried out with a third group of subjects, who were tested for reinstatement on Day 2 immediately after the extinction procedure.

The reinstatement test data showed a significant return of fear in response to CS+ category images in both the immediate and the 24 h reinstatement tests, making this study a non-replication of studies that abolished threat memory with the retrieval-extinction protocol. Return of fear was measured for CS+ and CS– separately, as the increase in SCR from the last trial of Day 2 extinction to the first trial of the post-reinstatement test. The authors state (p. 9), "We… demonstrated that a reminder before extinction did not prevent the recovery of generalized threat responses."

Evaluation of PE by the framework of Section 3 finds an absence of a PE experience in this study, which is the probable cause of the non-replication. Kroes et al. [132] provide extended and detailed discussion of the background research context as well as experimental design and results, but no mention of PE or mismatch is made, perhaps indicating an assumption that threat memory reactivation alone would be sufficient to induce memory destabilization, launching the MR process.

However, a PE experience is required for destabilization to occur, according to a preponderance of the empirical evidence, as reviewed in Section 3.1. Partial reinforcement during acquisition means that in this study, as in other retrieval-extinction studies with partial reinforcement, the single reminder CS+ presentation on Day 2 did not create a PE experience by being unreinforced, but it could create a PE experience by being single due to being followed unexpectedly by a nature video, mismatching the extended series of CS presentations learned in acquisition on Day 1. That PE experience would occur if, and only if, the transition after CS+ offset to the video appearing onscreen was perceived by subjects as entirely part of the study, not as a "break" happening outside of the study, thereby preserving the latent cause that had been active through CS+ offset, as explained in Section 3.3.1.

However, subjects in this study did perceive the coming video as a break happening outside the study because (p. 10) "During all sessions (acquisition, reminder, extinction, and recovery) with the exception of the breaks, SCR and shock electrodes were attached to the participants and the shock stimulator was set to the 'on' position." Disconnection from the electronic equipment for US shock delivery and SCR measurement certainly defined the video break as not being part of the study. Therefore, no PE experience was created by the Day 2 procedure, according to the PE evaluation framework of Section 3. Consequently, the category threat memory was not destabilized and could not be updated by the subsequent extinction training. This analysis is testable by repeating the study with SCR and shock electrodes remaining attached to subjects during the 10-min period and nothing communicated to subjects defining that period as a "break" from the study.

In contrast to the threat memory results, tests of episodic memory of the procedure showed that relative to the no-reminder group, the reminder group had significantly enhanced memory of items seen both before the reminder on Day 1 and after the reminder during extinction training. Enhancement has been observed as an effect of reconsolidation on episodic memory if no interference manipulation occurs [98, 146, 185]. In addition, the reminder group had "better item recognition specific to items from the CS+ category after 24 h but not immediately" (p. 4). The authors point out that that delayed effect is another distinctive feature of a reconsolidation process. To understand these results, the concept of PE specificity, introduced in Section 3.3.2, seems relevant. If indeed episodic memory underwent reconsolidation in this study, the reminder CS+ image must have created a PE experience specific to the episodic memory, even though it appears not to have created a PE experience specific to the CS+ category threat memory, as analyzed in the preceding paragraph. Those are two different memory systems, and in order to understand the results of this study, it may be fruitful to consider the possibility that the reminder CS+ image created a PE experience for the one memory system but not the other.

The obvious candidate for episodic memory PE is the novelty of the reminder CS+ image. That specific-image novelty would not constitute PE for the generalized CS+ threat memory because the reminder image is merely another member of that category, but the image does have novelty in relation to visual episodic memory. The fact that all CS+ images were novel during Day 1 acquisition does not eliminate the novelty of the Day 2 reminder image. Such visual novelty seems similar to the object recognition novelty that has been found to induce object memory reconsolidation [195]. Li et al. [113] similarly showed that a novel visual modification made in a familiar reminder image prevented return of fear in response to both CS+ and US presentations (both spontaneous recovery and reinstatement), whereas the unmodified reminder image prevented return of fear only in response to the US but not the CS+. The added image novelty evidently created its own visual image PE experience that destabilized the visual component of the threat memory in addition to the category recognition component (discussed more fully in Section A.12). This analysis of the episodic memory results of Kroes et al. [132] could be tested by repeating the study and adding another group of subjects for which the CS+ reminder image on Day 2 is not novel, but rather is a repetition of one of the more salient images used during Day 1 acquisition. Without the visual image mismatch, this group should not show the episodic memory enhancements if the above account is correct.

Thus, all results reported by Kroes et al. [132] appear to be consistent with and explainable by the PE evaluation framework defined in Section 3.

B.9. Fernandez-Rey et al. (2018) [184]: "Exploring the boundaries of post-retrieval extinction in healthy and anxious individuals"

Fernandez-Rey et al. [184] repeated experiment 1 of Schiller et al. [6] with several variations: The aversive US was a white noise burst instead of electric shock, a post-reactivation delay of 20 min was tested in addition to the standard 10-min delay with a "cartoon video" displayed onscreen in each case, and there was a 48-h instead of 24-h period between the memory manipulation procedure on Day 2 and testing for spontaneous recovery via re-extinction on Day 4. All other parameters duplicated those of Schiller et al., including measurement of fear via SCR.

The authors' Figures 2 and 3 as well as their statistical analyses show that comparing the CS+ SCR level on the last trial of Day 2 extinction to first trial of Day 4 re-extinction shows a significant return of fear in both retrieval-extinction groups (10-min delay and 20-min delay), and an even larger return of fear in the control group (no reactivation before extinction). In the first re-extinction trial in Figures 2 and 3, the CS+ SCR level returns to about 60% of peak acquisition level for the 10-min delay and about 80% of peak acquisition level for the 20-min delay. The authors acknowledge that aspect of their data by stating (p. 236), "the results of our comparison between the last extinction trial and the first re-extinction trial are not in keeping with the findings of [Schiller et al. [6]]." On that basis, this study is regarded in this review as a non-replication of human behavioral updating and annulment of emotional memory. (The authors mainly emphasize, however, that the CS+/CS– differential SCR level during the "early phase" of re-extinction, i.e., averaging the first several trials, shows no return of fear, and therefore they characterize their study's results as showing that (p. 230), "These findings suggest that postretrieval extinction is an effective behavioral technique for modifying the original fear memory and for the elimination of the fear return." They also comment (p. 236), "It is probable, then, that our results here confirm the suggestion…that different strategies used to analyze the extent of fear return can lead to inconsistent results.")

Also apparent in the authors' Figures 2 and 3 at the first re-extinction trial is a strong SCR level in response to the neutral CS– stimulus, about equal to the SCR for the CS+ in both retrieval-extinction groups. A very similar pattern of significant, equal SCR for both CS+ and CS– at the first re-extinction trial has been reported in [117] for one of their three tested reactivation conditions. That data can be explained as an effect of updating produced by the modified CS+ image that was used for reactivation in that condition, as reviewed in Section A.16 above. Fernandez-Rey et al. [184] did not purposefully modify the CS+ image in their study, but the analysis given for the modified image case introduces the possibility that a visual image PE experience could be created by any minor, unintended change of the image anywhere in the computer screen, such as a small screen artifact, a small label or logo newly appearing at the edge or corner of the screen, or a change of the background tone. A small visual change on the computer screen, or possibly even in the field of view outside of the computer screen, might be deemed unimportant by the researchers yet be sufficiently salient to function as a visual mismatch that creates the same type of visual PE experience, with the same effects, as described in Section A.16. It is probably more likely, however, that the heightened SCR amplitude in response to CS– was due to other causes of this effect that have been discussed in extinction and reconsolidation research literature, as discussed by Zuccolo and Hunziker [50], who also reported that effect.

The authors repeatedly refer to the Day 2 post-reactivation delay as a "rest period". If, immediately following CS+ reminder offset, subjects were informed explicitly that they would now have a "rest period" with a cartoon to watch, it is almost certain that subjects would then regard the coming rest period as not being part of the study, inducing a change of latent cause and preventing the PE experiences that would have occurred if no such communication had been made, in noted also in other reviews in Appendix B above and as explained in Section 3.3.1. Therefore absence of a PE experience is the probable cause of this non-replication, as indicated in Table 2.

B.10. Kredlow et al. (2018) [53]: "Exploring the boundaries of post-retrieval extinction in healthy and anxious individuals"

Kredlow et al. [53] used the retrieval-extinction protocol as in experiment 1 of Schiller et al. [6], with these parameter changes: Acquisition training on Day 1 had partial reinforcement of 60% (instead of 38%) pairing the CS+ (a colored shape, unchanged) on 6 of 10 presentations with electric shock US. For some subjects, 5 of the 6 shocks were accompanied by a simultaneous 1-s scream sound, delivered via headphones. Each CS presentation had 8-s duration (instead of 4 s), and the inter-trial interval was 11±1 (unchanged). Healthy subjects received that acquisition training on Day 1. For some healthy subjects, the US was the shock alone, for others it was the shock and the scream. All anxious subjects had the shock and the scream. One group of anxious subjects received acquisition training only on Day 1 and another group of anxious subjects received the same acquisition training also on Day 2 and Day 3, for a total of three such acquisition trainings.

On the day after completion of acquisition training (either Day 2 or Day 4), subjects underwent either PRE (post-retrieval extinction) or E (extinction). For PRE (p. 47), "the fear memory was retrieved with a single non-reinforced presentation of the CS+ (8 s duration). This was followed by a 10-min break during which participants watched a video about a sandcastle competition, which was chosen for its benign, non-emotional content. A series of standard extinction trials followed – participants were presented with 10 CS+ and 11 CS- trials non-reinforced." For E, subjects experienced only the video and the extinction training (11 of each CS, unreinforced and randomized).

On the day after PRE or E, fear memory responsiveness was tested via reinstatement by 4 unsignalled US shock presentations (3 of which were accompanied by the scream for subjects who had the scream during acquisition) spaced apart by 12 s. After 10 min, SCR was measured in response to 15 unreinforced presentations of each CS. "To test for reinstatement, we evaluated the change in average differential SCR from the last two trials of extinction to the first two trials of reinstatement."

The resulting data was surprisingly uninformative. The PRE groups had a moderate return of fear that was approximately equal to the return of fear in the E groups, and none of the particular experimental conditions created an effect distinct from the other conditions. The authors identify two possible causes of those results: (a) SCR for the neutral CS– was higher than SCR for CS+ toward the end of Day 2 extinction and was about equal to the elevated SCR for CS+ in the first trial after reinstatement. The authors regard this as indicating a strong generalization of fear that distorted the differential SCR data and also maintained elevated fear of the CS+ independently of any PRE-induced updating. (b) Selecting subjects who produced strongly differential SCR during acquisition may have resulted in a cohort with an unusually high level of contingency awareness, which is believed to reduce the effectiveness of the retrieval-extinction protocol, as discussed in Section B.1 above. Those two effects have been observed in other studies, so attributing this study's non-replication results to them has stronger plausibility than mere speculation. Both effects are artifacts of experimental design, not fundamental weaknesses of the behavioral updating mechanism.

The authors did not also address the possible role of PE in their results. As in other non-replication studies with partial reinforcement, the CS+ reminder presentation did not create a PE experience by being unreinforced, so whether any PE occurred depended on how the transition to the 10-min video was made, as explained in Section 3.3. The authors state (p. 47), "all participants were attached to the shock electrodes and wore headphones throughout the entire experiment." That eliminates removal of equipment as one way subjects could receive the message that the video is not part of the study, changing the active latent cause and preventing the video from generating a PE experience. However, referring explicitly to the imminent video as a "10-minute break" would also have that effect, and it is not indicated whether that communication was made. Some subjects may have interpreted the video content, a sandcastle competition, as meaning that this period was not part of the study. Therefore an absence of any PE experience is an open possibility as the cause of this study's non-replication.

B.11. Zuccolo and Hunziger (2018) [50]: "Preliminary study of the effects of post-retrieval extinction on the return of conditioned responses in humans"

These researchers attempted the important task of exact replication of results of experiment 2 of Schiller et al. [6], varying the procedure only by testing for return of fear on Day 3 using reinstatement rather than re-extinction. However, replication was not achieved, and the authors report (p. 131) that "results did not replicate the differential effects described by Schiller et al. (2010). …[I]n the first trial [after reinstatement], participants from both groups showed an increase in responses to all CSs+."

Zuccolo and Hunziker [50] provide a detailed account of their procedure for the transition from Day 2 CSa+ reminder offset to the start of the 10-min delay period. According to the trial PE evaluation framework proposed on Section 3, that brief element of procedure pivotally determines whether destabilization and updating will occur in studies that used partial reinforcement in acquisition training, but it has rarely been documented in study reports and even more rarely included in discussions of non-replications. The authors explain (as noted in Section 3.3.1 above but repeated here necessarily) (pp. 129–130):

After a single presentation of the [reminder CS], the computer screen went black, and the following instructions were given: "We now will have a ten-minute break. Here are some magazines that you may choose to read during the break. You do not have to read anything if you do not want to. The only thing that is important is that you take your break in the specified location where I take you. However, first, I will turn off and disconnect you from these devices". The subject was disconnected from the electrodes, the stimulator was set to the "Off" position, and the participant was taken to a waiting room with chairs and magazines. …Once the ten-minute period was over, participants were returned to the testing room, and the following instructions were given: "We have finished the ten-minute break. I will now reconnect these devices, and we will complete the remainder of the session". Subjects were reconnected to the electrodes, and the stimulator was set to the "On" position. The session was resumed with the remaining presentations of the CSs (i.e., 14 CSa+, 15CSb+, and 14 CS-).

With that transition, it can be inferred reliably that immediately after CSa+ offset, subjects regarded the coming "break" as occurring outside of the study and that they therefore construed a different active latent cause than was just previously in effect during the study's CSa+ reminder presentation. The CSa+ threat memory was then no longer the active expectational framework, so now even unexpected presentations could not create a PE experience, because now there was no memory-based expectation in effect. The 10-min period and its contents were now subjectively irrelevant to the CS+ memory and its latent cause, so the magazines offered did not have the meaning of PE in relation to the target CSa+ threat memory. That memory was therefore not destabilized, resulting in it being merely suppressed temporarily rather than transformed through updating by the extinction procedure to follow. There is also no PE experience created by the appearance of only one CS before the break because the expected series of CSs was perceived as being interrupted by events having a different latent cause than the CS images, so the interruption is not construed as the CSa+ memory's PE.

That analysis predicts that updating and nullification of the CSa+ threat memory would occur if this study were repeated with no communication to subjects about the transition and the transition carried out in either of these two different ways, in two groups of subjects: (a) CS+ reminder offset is followed by a 2-min inter-trial interval and then extinction training begins. The longer-than-expected time interval should create a PE experience [109, 111]. (b) CS+ reminder offset is followed by the expected inter-trial interval of 11±1 s, at which point the computer screen displays a nature video for 2 min, and then extinction training begins. The unexpected appearance of the video instead of the expected next CS image should create a PE experience. Each of those procedures, if successful at producing updating and preventing return of fear, would prove that what is inherently necessary for reconsolidation updating is not a 10-min delay, but rather a PE experience produced by any suitable procedure.

B.12. Chalkia et al. (2020a) [52]: "No persistent attenuation of fear memories in humans: A registered replication of the reactivation-extinction effect"

Here again the important task of exactly replicating the seminal study by Schiller et al. [6] was undertaken. Chalkia et al. [52] pre-registered their study and assembled a larger sample size and stronger statistical power than the original study or any previous test of the retrieval-extinction protocol. However, they reported (p. 504), "despite following the protocol of Schiller et al. (2010) in detail, we observed spontaneous recovery of fear and susceptibility to reinstatement in both the reactivation and no reactivation groups, without any significant between-group differences."

In attempting to explain this striking failure of replication, the authors focus mainly on the high levels of participant exclusion in various successful human retrieval-extinction studies, including Schiller et al. [6]. Specifically, they suggest that (pp. 506–507) "With such rates of participant exclusions, the resulting sample may not be representative of the general population but rather of a specific subset of individuals that have a particular aptitude at learning from sparse contingency information. …That is to say, not everyone learns readily under such a low reinforcement rate [38%], implying that the participants that were included may be somewhat unique in their learning capabilities. Perhaps, one might be tempted to argue, reactivation-extinction is indeed beneficial for this selected sub-group. …It may be the case that unknown individual differences moderate reactivation-extinction effects." However, they add (p. 506), "Yet, if reactivation-extinction was indeed advantageous for certain sub-groups of participants only, we should have been able to observe this benefit in the present study, considering that we used the same partial reinforcement schedule and similar inclusion criteria as in Schiller et al. (2010)." Thus, they do not arrive at a consistent, probable explanation.

No consideration is given by Chalkia et al. [52] to the possible role of PE. Applying the current review's trial framework of PE evaluation, as applied to the retrieval-extinction protocol (Section 3.3.1), points to absence of a PE experience as a strong candidate explanation, as follows.

The single unreinforced CS+ presentation on Day 2 would normally create a PE experience, though not by being unreinforced, due to partial reinforcement during acquisition on Day 1, but rather by occurring alone, without being followed by other CSs every 11±1 s as expected. In addition, seeing a TV show appear next, instead of the expected next CS image, would also create a PE experience, ensuring prompt destabilization and later updating of the CS+ threat memory.

However, the actual transition from the CS+ reminder to the TV show was done by Chalkia et al. [52] in a manner that prevented both of those PE experiences, as analyzed by the trial PE framework. In describing the Day 2 procedure, Chalkia et al. explain (p. 500), "Participants in the reactivation group…received a single, unreinforced CS+ trial, followed by a 10-min break…, during which all participants watched a preselected Simpsons episode….All electrodes remained attached (Schiller et al., 2010), but participants were informed explicitly that no shock would be administered during the break [italics added]. After the break, both groups proceeded immediately to extinction training." Thus, immediately after the CS+ reminder they had just observed, the subject without delay received a communication from the researchers implying strongly that what was now happening was an entertaining "break" that was not part of the study. The subject was now regarding the TV show as not being part of the study before it began playing, so it did not have the same latent cause as the CS+ reminder they had just observed, was therefore subjectively not relevant to the CS+ and its latent cause, and therefore did not create a PE experience in relation to the target CS+ threat memory, which therefore was not destabilized, precluding later updating.

Thus, failure to create a PE experience has to be considered a viable candidate explanation for the replication failure of Chalkia et al. [52]. This explanation could be tested by conducting the replication again without any communication from researchers regarding the post-reactivation transition, and now with CS+ reminder offset followed by the expected inter-trial interval of 11±1 s and then by the TV show, with TV show duration of 10 min for one group of subjects and durations of 30 s and 2 min for two other groups. For at least the 10-min group, that should allow both of the PE experiences described above to occur, destabilizing the CS+ threat memory and allowing it to be updated and nullified by the subsequent extinction training. If updating also occurs for either of the shorter TV show durations, that would prove that what is inherently necessary for reconsolidation updating is not a 10-min delay, but rather a PE experience produced by any suitable procedure.

B.13. Houtekamer et al. (2020) [129]: "Investigating the efficacy of the reminder-extinction procedure to disrupt contextual threat memories in humans using immersive Virtual Reality"

The conclusion drawn by Houtekamer et al. [129] from this study of the retrieval-extinction protocol applied to human contextual fear memory was (p. 12), "we found no evidence that the reminder-extinction procedure is a more effective procedure to modify contextual threat conditioned memories in humans as compared to regular extinction."

The absence of any PE experience in the procedure explains that replication failure. The Day 1 acquisition training consisted of a 60% reinforcement schedule, with subjects receiving an electric shock in 6 of 10 visits to the onscreen conditioned context (CTX+). Subjects in the retrieval-extinction group experienced on Day 2 a reminder task consisting of a single no-shock visit to the CTX+. That experience reactivated the target memory but did not create a PE experience because a single visit to CTX+ with no shock is fully consistent with the target memory. Therefore, the reminder did not destabilize the target memory, contrary to the researchers' apparent assumption that memory reactivation alone would do so.

The authors state (p. 6), "In line with previous studies, the reminder was followed by a 10 min break, so that the following extinction task would fall within the putative reconsolidation window. During this break, all participants (both the R-Ext as Ext groups) watched 10 min of landscape scenes from BBC Planet Earth (2006 TV series). Participants were explicitly told that they would not receive any shocks during this break, and the shock equipment was visibly turned off for the duration of the break. All participants were then told that the task would continue as before, that they would again hear sounds and might receive shocks."

With that transition to the video, subjects definitely regarded the video as an event fully outside the study, inducing a change of latent cause as discussed in Sections 3.3.1 and 5. The target memory was therefore not destabilized and the subsequent extinction training functioned as standard extinction, a condition no different than for the standard extinction group of subjects. Consequently, on Day 3 no significant difference in return of fear was found between the two groups, and the study was a replication failure. It is instructive to contrast that transition with the transition carried out by Kitamura et al. [128], reviewed in Section A.17, which caused subjects to view the coming video as being an integral part of the study, preserving the latent cause that was already active and allowing the ensuing video to produce a PE experience that destabilized the target memory, resulting in updating and no return of fear.

If Houtekamer et al. [129] had used 100% reinforcement at acquisition, the same Day 2 procedure for the retrieval-extinction group would have created a memory mismatch and a PE experience due to the single no-shock visit to the CTX+, which would have produced destabilization and emotional memory annulment, by this analysis. Testing that prediction could provide a confirmation of this analysis.

The authors do consider the PE requirement toward the end of an extended discussion of possible causes of their replication failure (p. 13): "Alternatively, …our reminder procedure may have failed to reactivate memory in such a way that it resulted in destabilization of the memory. If indeed the efficacy of the reminder-extinction procedure depends [on] memory destabilization and disruption of a reconsolidation process, generation of a prediction error during the reminder may be critical for successful destabilization."

B.14. Zimmermann and Bach (2020) [197]: "Impact of a reminder/extinction procedure on threat-conditioned pupil size and skin conductance responses"

Once again, replicating experiment 2 of Schiller et al. [6] was the purpose of this study, but with SCR measurement supplemented by measurement of another autonomic response, pupil size response (PSR). The authors note that (p. 165) "a direct comparison has revealed that PSR may have higher accuracy in inferring fear memory than SCR (Korn et al. 2017)." These researchers also (p. 165) "optimized some parameters in light of a recent meta-analysis by Kredlow et al. (2016), in order to increase the chances of revealing success of the reminder/extinction procedure…." The parameters used by Zimmermann and Bach [197] are compared to those of [Schiller et al.] in this list:

- Conditioned stimuli (CS): colored triangles [colored squares]
- CS presentation: 4 s [4 s]
- Intertrial interval: 9±2 s [11±1 s]
- Electric shock (US): 500 msec, 500 Hz [200 msec, 50 Hz]
- Reinforcement rate: 50% [38%]
- Acquisition: 16 CSa+/CSb+, 10 CS– [13 CSa+/CSb+, 8 CS–]
- Reminder: CSa+ [CSa+ and CS–]
- Extinction: 10 CSa+, 11 CSb+/CS– [10 CSa+, 11 CSb+/CS–]
- Reinstatement: 4 unsignalled US [4 unsignalled US]
- Reextinction: 10 CSa+/CSb+/CS– [10 CSa+, 11 CSb+/CS–]

In addition to those optimizations, this study's statistical power was slightly stronger than that of Schiller et al. [6]: "Our sample size was based on the signal-to-noise ratio of PSR under control conditions. Thus, we recruited a sample that provided 85% power to detect an at least 50% absolute reduction of fear retention, corresponding to the ~60% reduction found in Schiller et al. (2010)" (p. 165).

Nevertheless (p. 168), "we found no evidence that a reminder trial before extinction prevents the return of fear in several analyses and both outcome measures. When comparing the last trial of extinction with the first trial after reinstatement for each condition, we observed significant reinstatement equally for the reminded and the nonreminded CS+…. To exclude that our negative findings are due to the inclusion of participants that did not successfully learn or extinguish the CS/US associations, we performed a supporting analysis with the same exclusion criteria as Schiller et al. (2010) (published in Schiller et al. (2018)), thus including N = 50 participants for PSR and N=22 participants for SCR. This should provide similar or higher power than Schiller et al. (2010) (N=18 participants). After applying these criteria, we did not observe any difference between reminded and nonreminded CS+ either. To summarize, we find no evidence that a reminder/extinction procedure prevents return of fear in our study."

Analyzing this study's non-replication by the PE evaluation framework of Section 3 requires examining the transition from Day 2 reminder CS+ offset to the 10-min post-reminder delay period. The authors provide these details (p. 170): "A 10 min break separated the reminder trial from the extinction session, during which the participants stayed attached to the recording electrodes but were explicitly instructed that no shock would be delivered [italics added]. During the break all participants watched a preselected TV show without audio but subtitles." That is identical to the transition procedure described by Chalkia et al. [52], whose high-powered study also was a non-replication, as reviewed in Section B.12. Therefore the non-replication by Zimmermann and Bach [197] has the same probable cause, namely, absence of PE experience due to an immediate change of active latent cause after CSa+ reminder offset, induced by the researchers' communication to the subject. This analysis could be tested in the same way as indicated at the end of Section B.12.