### The lack of evidence in Schiller et al. (2010) verified: Reply to Schiller, LeDoux, and Phelps (2020)

Anastasia Chalkia<sup>a,b</sup>, Lukas Van Oudenhove<sup>b,c,d</sup>, and Tom Beckers<sup>a,b</sup>

<sup>a</sup> Centre for the Psychology of Learning and Experimental Psychopathology, Faculty of Psychology & Educational Sciences, KU Leuven, Leuven, Belgium

Schiller, LeDoux, and Phelps recently posted a preprint (Schiller et al., 2020) in which they comment on a Verification Report we published (Chalkia et al., 2020a; VR) regarding their initial demonstration of the retrieval-extinction effect in humans (Schiller et al., 2010). The contents of the preprint are also partly concerned with a Registered Report we published regarding a direct replication of the same study (Chalkia et al., 2020b; RR). They also comment on the editorial handling of our VR (details of which can be found in McIntosh & Chambers, 2020) and the general purpose of VRs (an introduction to which can be found in Chambers, 2020).

Of note, their reply fails to counter, or even acknowledge, the most crucial finding in the VR – the fact that the original results of Schiller et al. (2010) were entirely dependent on arbitrary participant exclusions. The reported exclusion criteria for their Experiment 1 and Experiment 2, despite changing with every report (see Schiller et al., 2010, 2012, 2018), were never actually applied. In the VR, we demonstrated empirically that apart from the allegedly *qualitative* selection of specific participants by Schiller et al. (2010) into the final datasets, any other combination of included participants failed to yield evidence for the effect these authors claimed to have obtained. Interestingly, through this qualitative selection by the original authors, participants without any evidence of fear learning and/or extinction learning were included in the final datasets, which according to Schiller et al. (2020, p. 2) is poor practice because in the absence of learning, there can be no memory to retrieve.

The analyses included in the VR clearly document this critical discovery and we will not reiterate them here. We will also not respond to the lengthy defense in Schiller et al. (2020) of the phenomenon of reconsolidation or the veracity of the retrieval-extinction effect, given that we never generalized the observations in our VR regarding the findings of Schiller et al. (2010) to the wider literature. Lastly, we will not address the comments critiquing the editorial process or questioning the purpose of VRs, as we feel in no position to do so. If Schiller et al. (2020) have their own suggestions about how a VR should be conducted and handled, they can bring them to the attention of the appropriate people (the aims and policies

<sup>&</sup>lt;sup>b</sup> Leuven Brain Institute, KU Leuven, Leuven, Belgium

<sup>&</sup>lt;sup>c</sup> Laboratory for Brain-Gut Axis Studies (LaBGAS), Translational Research Centre for Gastrointestinal Disorders (TARGID), Department of Chronic Diseases and Metabolism, KU Leuven, Leuven, Belgium

<sup>&</sup>lt;sup>d</sup> Cognitive and Affective Neuroscience Lab, Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

of VRs can be found online here:

https://www.elsevier.com/ data/promis misc/VR GuideForAuthors.pdf).

What we will do in our rebuttal is respond to the criticisms raised by Schiller et al. (2020) regarding the contents of our VR, which these authors tried to dismiss as an opinion piece rather than the factual report that it is. Below, we repeat (in bold) the claims that Schiller et al. (2020) argue we make in our VR, followed (in italic bold) with their criticism of each claim. We show that their arguments are either misguided or incorrect, and, most importantly, we point out, once again, that their original results were entirely dependent on selective participant exclusions (see points 2, 3, and 4 below). Further, while we never previously referred to their participant exclusions as post-hoc, below, we now detail the evidence we have that demonstrates the post-hoc nature of their participant exclusions which undeniably influenced the results (see point 3 below). In sum, we maintain that the results reported by Schiller et al. (2010) are critically flawed.

# 1) Replication attempts of the original study are few and have yielded mixed results. To date, there have been at least 51 published replication studies of retrieval-extinction in humans and clinical populations, with over 70% success.

In their Table 1, Schiller et al. (2020) list a number of replication attempts that we did not mention in our VR and RR. In our VR, we actually stated that "a multitude of conceptual replication studies have been published, some successful and others not" (Chalkia et al., 2020a, p. 511), while in our RR we stated that "attempts at conceptual replication of these findings have yielded mixed results" (Chalkia et al., 2020b, p. 497). Although Schiller et al. (2020) claim that we mentioned just a few replication attempts in our VR and RR, in our RR we acknowledged almost all of the replication studies listed in Schiller et al. (2020, Table 1) that were published by the time our manuscript was submitted for peer review (our Stage 1 RR manuscript was submitted in 2016). According to the policies of registered reports, authors are not allowed to change their introduction and methods section after acceptance of their Stage 1 manuscript. However, we did add a few more published studies as footnotes (both successful and failed replications) in our Stage 2 RR. In our VR, we merely pointed our readers to published reviews (i.e., Auber, Tedesco, Jones, Monfils, & Chiamulera, 2013; Monfils & Holmes, 2018; Zuccolo & Hunziker, 2019) and a meta-analysis on the retrievalextinction effect (Kredlow, Unger, & Otto, 2016). As such, we think our portrayal of the published literature was balanced. The purpose of a VR is not to comprehensively review the literature, but to test the reproducibility and analytical robustness of one specific published result, which is exactly what we did.

## 2) The verification report was unable to verify the results reported in the original study. The verification report was in fact able to reproduce the results almost entirely.

Schiller et al. (2020) claim that our VR was able to reproduce the original results almost entirely, and that therefore its conclusions do not reflect the data, but rather, our opinion. Those who have read our VR in any detail will know that reproducing the original results was far from straightforward. A lot of investigative work went into figuring out how exactly Schiller et al. (2010) had gone about processing their data, calculating their variables, and

designing their analyses. Moreover, we had to establish which participants they had excluded from those analyses, because of discrepancies between their report and their actual analyses. The authors in reality tested 77% more participants (N = 126) than they originally reported (N = 71) and excluded 48% rather than 8% of participants in Experiment 1. For Experiment 2, these rates were even worse, as they tested 233% more participants (N = 70) than they originally reported (N = 21) and actually excluded 74% rather than 14% of their sample. We will not reiterate all those issues in full detail here, as they can be found in the VR (Chalkia et al., 2020a).

When we eventually managed to reproduce their results, we found that those results provided little statistical support for Schiller et al.'s (2010) central claims, as critical interactions testing their hypotheses were not significant. However, as we mentioned above, the major claim in our VR, which Schiller et al. (2020) fail to comment on, is the fact that even those results could only be obtained when participants were carefully selected into the final datasets, in a way that cannot be justified even after the facts (see below). This statement does not reflect an opinion. We clearly demonstrate, empirically, that when any combination of rigorously applied exclusion criteria, including the different versions reported by the authors, was used instead, there is not even partial support for the retrieval-extinction effect in Schiller et al.'s (2010) data. Ultimately, not only was it challenging to reproduce the original analyses, due to Schiller et al.'s (2010) misreporting, but any of the further (more principled) analyses we conducted also failed to yield evidence favoring Schiller et al.'s (2010) hypothesis, demonstrating that the original results, even if they had been valid, were not analytically robust.

### 3) Exclusions made post-hoc influenced the results.

Most participants that did not show measurable responses or evidence for learning did not continue to the next day - a process that eliminates post-hoc exclusions.

According to Schiller et al. (2020), we unjustly claimed that post-hoc participant exclusions critically influenced the results. This is a somewhat peculiar argument, because we never claimed in our VR that participant exclusions in Schiller et al. (2010) were made post-hoc (but see below). We merely referred to the exclusions as qualitative, in accordance with the wording used in Schiller et al. (2018). In their reply, Schiller et al. (2020) insist that participants were only ever excluded (mostly prior to completing the full study) because of a lack of learning and/or for non-responding. This makes one wonder why some participants that did not show any evidence of learning were nonetheless allowed to complete the full study and subsequently even included in the final dataset, as we illustrated in the VR (see for example, Figure 3 in Chalkia et al., 2020a, p. 512). Schiller et al. (2020, p. 2) further claim: "The original article only reported exclusions from participants with complete datasets and misstated the exclusion criteria, which led us to publish an addendum to the original study (Schiller et al., 2018)." This is a highly misleading statement, for multiple reasons. First, the original article reported only 6 excluded participants for Experiment 1, while the OSF datafile contains the complete datasets of 19 excluded participants from Experiment 1. Second, regarding the misstated exclusion criteria, we leave it to the reader to decide whether the

change from 2 very strict exclusion criteria in Schiller et al. (2010) to a set of 8 conditional statements in Schiller et al. (2018) can be reasonably considered a *misstatement*.

Even so, these conditional statements do not comprehensively capture the reasons for the inclusion or exclusion of participants in the final dataset, as the authors acknowledged in Schiller et al. (2018), mentioning for example that 3 participants who met at least one of the acquisition criteria for Experiment 1 were excluded "for [displaying] idiosyncratic responses not observed in other participants in the sample (e.g., conditioned response on last acquisition trial was negative or almost equal to zero)" (Schiller et al., 2018, p. 2). However, as we reported in our VR (Chalkia et al., 2020a, p. 513), when inspecting the final dataset, one can observe that 8 included participants also displayed a negative conditioned response on the last acquisition trial, and another 11 included participants exhibited responding equal or almost equal to zero. It is entirely unclear why those [19] participants were then not also excluded for the same reason of idiosyncratic responding. It appears to be impossible to provide a conclusive characterization of the exclusions made by Schiller et al. (2010), even after the facts. Schiller et al. (2020) neglect to comment on this observation altogether.

Finally, we should revisit the alleged post-hoc nature of some of the exclusions made by Schiller et al. (2010). In our VR, we deliberately refrained from commenting on the process underlying participant exclusions in Schiller et al. (2010). However, given that Schiller et al. (2020) themselves now raise (and deny) the post-hoc nature of participant exclusions, we feel obliged to full disclosure. When we requested the original data underlying Schiller et al. (2010) from the report's lead author, Daniela Schiller (see Chalkia et al., 2020a; Chalkia et al., 2020b for details), she initially did not manage to comply, citing logistical difficulties (i.e., traveling, grant deadlines). After five months, the report's senior author, Elizabeth Phelps, arranged for her lab manager to send us all the archived files she had retained related to the study. Those files provide clear cues that at least some of the participant exclusions were post-hoc. Consecutive excel sheets, compiled while data collection for Experiment 1 was ongoing, contain paired t-tests (one-tailed and two-tailed) comparing the end of extinction to the beginning of re-extinction within each group and unpaired t-tests comparing each combination of two groups at the end of extinction as well as at the beginning of re-extinction, in different subsets of participants, both for CS+/CSresponding and for the differential SCR response. For the 1-year follow-up to Experiment 1, different combinations of extinction trials (e.g., last 2 OR last 4) are compared to postreinstatement responding or included for the calculation of separate reinstatement indices (i.e., post-reinstatement responding minus responding during extinction). These indices, and post-reinstatement responding per se, are then compared, across groups, using one-tailed and two-tailed *t*-tests. <sup>1</sup>

The chronological order of these excel sheets can easily be deciphered by comparing the data included in each excel sheet to the data included in the final dataset (for Experiment

<sup>1</sup> The previous version of this preprint contained a small error in the description of exactly which tests were performed in the intermediate analyses.

4

1). In an earlier version of those excel sheets, which contains both participants that were eventually included and excluded from the final dataset, all these analyses yield non-significant effects. However, in later versions, some participants have *disappeared* from the sample, from all groups, and the same analyses have been repeated. [Of note, these disappearing participants are the 19 excluded participants for which full data are contained in the OSF datafile, while the rest of the participants are the ones retained in the final dataset.] In that later version, the within-groups effects are in line with the authors' hypothesis, in that they show a lack of spontaneous recovery in the retrieval-extinction group and a recovery of fear in the control groups. The selection of participants into the final dataset after the conducting of intermediate statistical analyses clearly suggests the post-hoc nature of participant exclusions in the original study.

### 4) Experimenter decisions regarding exclusion cut-off influenced the retrievalextinction effect.

The retrieval-extinction effect remained intact under all exclusion cut-offs assessed in the verification report.

Schiller et al. (2020) claim that the VR merely showed that control groups did not show spontaneous recovery in any of the re-analyses. However, we can hardly be held accountable for the fact that none of the groups tested by Schiller et al. (2010) exhibited spontaneous recovery to begin with, unless specific participants (that actually showed a lack of spontaneous recovery) were excluded from the control groups, in a post-hoc fashion (see above). Of note, when using either the exact sample included by Schiller et al. (2010) or the full sample of participants available in the OSF dataset, we did observe significant spontaneous recovery overall, without significant between-group differences, even if withingroup comparisons failed to yield evidence for spontaneous recovery in the control groups (see Chalkia et al., 2020a, pp. 516 - 520).

### 5) Successful replication of the retrieval-extinction effect depends on high exclusion rates.

Studies that have successfully replicated the retrieval-extinction effect have participant exclusion rates ranging from 0% to over 50%.

Schiller et al. (2020) argue that successful replication of the retrieval-extinction effect does not depend on exclusion criteria or rates. We take no issue with this argument and agree that successful conceptual replications have been obtained in studies with limited or no exclusions. But for Schiller et al. (2010), clearly, high exclusion rates and idiosyncratic exclusion criteria were *essential* for obtaining their results.

#### 6) Statistical tests of the recovery effect were inappropriate.

An opinion about an alternative analysis is not the same as an error with the original analysis.

Schiller et al. (2020) argue that our claims regarding the appropriateness of the statistical tests for the retrieval-extinction effect (in Experiment 1) are a matter of opinion. Two separate issues are at play here. First, in our VR, we question the practice of performing multiple, uncorrected, within-groups *t*-tests in the absence of a non-significant group interaction effect

throughout Schiller et al. (2010) and not just for the spontaneous recovery comparison: "While some suggest that it is acceptable to perform multiple comparisons without an overarching significant ANOVA interaction, it is generally recommended to then correct for inflation of the type-I error rate through procedures such as Bonferroni adjustment or others" (Chalkia et al., 2020a, p. 522). One may indeed regard this as a matter of opinion (although then not a very controversial opinion). The other issue is that when Schiller et al. (2010) followed up their spontaneous recovery ANOVA with separate *t*-tests per group, they used different time points for their *t*-tests than those used in the ANOVA – a fact that can be verified simply by reading their manuscript. They followed a similar approach for the reinstatement test in Experiment 2, whereas for the long-term reinstatement test in Experiment 1 they did use the same time points for the ANOVA and the follow-up *t*-tests. To chalk this inconsistency up to a matter of opinion is simply disingenuous.

One can argue about whether differences in the degree of re-extinction (within-phase) or differences in the degree of recovery from the end of extinction to the recovery test (between phases) represent the most appropriate analysis to test for differences in memory retention (we take no position in that here), but clearly the follow-up paired t-tests should relate to the same data points as the ANOVAs they follow up on. As we report (Chalkia et al., 2020a, p. 516, p. 520, p. 522), when ANOVAs are performed that match the ensuing t-tests, all critical interactions testing the retrieval-extinction effect become clearly non-significant (Experiment 1: p = .091 for differences in spontaneous recovery between groups, p = .114 for differences in long-term reinstatement between groups; Experiment 2: p = .34 for differences in reinstatement between cues). Schiller et al. (2020) denote these as trends that in the presence of clear a priori directional hypotheses merit follow-up testing, which at best might be true for the first one. However, even obtaining this trend depends on very idiosyncratic participant selection, as pointed out above.

Going further into the spontaneous recovery ANOVA, we also questioned Schiller et al.'s (2010) choice to use only part of the re-extinction phase data, without justification. While the re-extinction phase consisted of 10 trials, they only included the first 8 trials in the ANOVA: "Spontaneous recovery was assessed using a two-way ANOVA with main effects of group (10 min, 6 h, and no-reminder) and time (early and late phase of re-extinction, defined by the mean first 4 responses versus the subsequent 4, respectively)" (Schiller et al., 2010, p. 50). This is unlike other analyses in their report, where the data from complete phases were typically used (i.e., 10 trials for acquisition and extinction). To gauge the impact of this choice, we here take the liberty of re-executing the spontaneous recovery ANOVA as preferred by Schiller et al. (2010), evaluating differential responding between groups across Day 3, and including either the first 8 trials only or all 10 trials. When only the first 8 trials of the re-extinction phase are included in the analysis (comparing the first 4 to the next 4 trials), we are able to exactly replicate the analysis reported by Schiller et al. (2010), showing a significant main effect of time, F(1, 62) = 6.27, p = .015,  $\eta_p^2 = .09$ , and a significant group by time interaction, F(2, 62) = 4.63, p = .013,  $\eta_p^2 = .13$  (as reported in Chalkia et al., 2020a, p. 516). However, when the complete re-extinction phase is used instead (comparing the first 5 to the last 5 trials, analogous to the other analyses in the paper), we find evidence for a decline in fear responding over time, across groups, without significant between-group differences (main effect of time, F(1, 62) = 4.13, p = .046,  $\eta_p^2 = .06$ ; group × time, F(2, 62) = 1.11, p = .34,  $\eta_p^2 = .04$ ). The choice of trials to be included in the analysis, while seemingly erratic, thus appears to have been important for obtaining a significant interaction effect in the ANOVA that was supposed to provide the basis for the ensuing paired-samples *t*-tests.

### 7) Failure to reproduce critical evidence for reinstatement in study 2.

The analysis reported in the verification report fully supports reinstatement in study 2. Schiller et al. (2020) correctly point out that the analysis testing for reinstatement in Experiment 2 presented in the VR was different than what was reported in Schiller et al. (2010). Whereas in Schiller et al. (2010) the reinstatement analysis involved a comparison of stimuli (CSa+, CSb+, CS-) across time (early versus late phase of re-extinction, defined as the mean of the first four versus the last four trials, respectively), our verification analysis compared the same stimuli (CSa+, CSb+, CS-) across slightly different time points (early versus late phase of re-extinction, defined as the mean of the first four versus the next four trials, respectively). We unjustly assumed that Schiller et al. (2010) maintained a similar approach for their reinstatement comparison (Experiment 2) as for their spontaneous recovery comparison (Experiment 1). A correction to our VR has meanwhile been published to rectify this (Chalkia et al., 2021; see below).

That being said, it is entirely unclear why Schiller et al. (2020) maintain that in our (erroneous) verification analysis we reported a main effect of stimulus, which they now insist is the critical test for reinstatement and thus in support of their hypothesis. In the VR one can read: "With respect to the recovery of fear, unlike what Schiller et al. (2010) reported, we found that a  $3 \times 2$  ANOVA comparing the first 4 to the subsequent 4 trials of re-extinction revealed no significant effects (main effect of stimulus, F(2,34) = 2.82, p = .074,  $\eta_p^2 = .14$ ; main effect of time, F(1,17) = 2.83, p = .11,  $\eta_P^2 = .14$ ; stimulus × time, F(1.14), 19.33) = 1.01, p = .34,  $\eta_p^2 = .06$ )." Schiller et al. (2010) did report a significant stimulus by time interaction (F(2,51) = 5.14, p < .01), using slightly different trials, but the degrees of freedom in their ANOVAs for this experiment are inconsistent with the stated analytic design and sample size (as noted in the VR and not commented on by Schiller et al., 2020). When we conduct the exact same analysis as reported in Schiller et al. (2010), i.e., a  $3 \times 2$  ANOVA comparing the first 4 to the last 4 trials of re-extinction, we now do find a main effect of stimulus (F(2,34) = 3.98, p = .028,  $\eta_p^2 = .19$ ), but no main effect of time nor interaction (main effect of time, F(1,17) = 0.33, p = .57,  $\eta_p^2 = .02$ ; stimulus × time, F(2,34) = 1.15, p = .33,  $\eta_{\rm p}^2 = .06$ ) (Chalkia et al., 2021). So, even when analyzing the trials as reported by Schiller et al. (2010), we are unable to reproduce their reported results; we do however find a main effect of stimulus (which neither Schiller et al., 2010, nor Chalkia et al., 2020a, had reported).

Schiller et al. (2020) now maintain that a main effect of stimulus [which they never reported before] is the critical result. They even state that "The stimulus  $\times$  time interaction [which we did not reproduce] only adds that the difference between the stimuli decreased over time during re-extinction, which is *not* the critical test of reinstatement" (Schiller et al.,

2020, p. 3). We simply note that they did report this interaction (and not the main effect) as the critical test of reinstatement in Schiller et al. (2010).

### To sum up: Major issues raised in the VR were disregarded by Schiller et al. (2020)

Schiller et al. (2020) claim to respond to the major issues raised in our VR, but they actually fail to address the most critical ones. To highlight just a few of them: 1) The pattern of results reported by Schiller et al. (2010; across all experiments) can only be obtained when participants are hand-picked into the final datasets, without the rigorous application of any sort of criteria. Moreover, as we describe above, statistical analyses were conducted prior to the *qualitative* selection of participants into the final datasets; 2) The results of Schiller et al. (2010, Experiment 1 – Reinstatement) can only be obtained if a different phase of the experiment is included in the analysis than what was reported and different trials are excluded for the experimental group than for the control group (augmenting the differences between the groups in the intended directions); 3) The results of Schiller et al. (2010, Experiment 2) are a conundrum – neither we nor they seem to be able to computationally reproduce them; 4) The number of participants originally recruited was not disclosed until 8 years after the publication of Schiller et al. (2010) and still remains unreliable (i.e., Schiller et al., 2018 report the recruitment of 126 participants for their Experiment 1, 61 of which were excluded, and go on to describe exclusions that add up to 66).

The list of issues goes on, but we will stop here. Schiller et al. (2020) may well wish that our VR was merely an opinion piece, but we presented just the facts. And so, while we agree that much remains to be discovered about the scope and mechanisms of the retrieval-extinction effect, in humans and other animals, we stand by our evaluation that Schiller et al. (2010) makes no valid contribution to this literature and should not be considered when weighing the evidence base for this effect.

#### References

- Auber, A., Tedesco, V., Jones, C. E., Monfils, M. H., & Chiamulera, C. (2013). Post-retrieval extinction as reconsolidation interference: Methodological issues or boundary conditions? *Psychopharmacology*, *226*(4), 631–647. https://doi.org/10.1007/s00213-013-3004-1
- Chalkia, A., Schroyens, N., Leng, L., Vanhasbroeck, N., Zenses, A.-K., Van Oudenhove, L., & Beckers, T. (2020b). No persistent attenuation of fear memories in humans: A registered replication of the reactivation-extinction effect. *Cortex*, *129*, 496–509. https://doi.org/10.1016/j.cortex.2020.04.017
- Chalkia, A., Van Oudenhove, L., & Beckers, T. (2020a). Preventing the return of fear in humans using reconsolidation update mechanisms: A verification report of Schiller et al. (2010). *Cortex*, 129, 510–525. https://doi.org/10.1016/j.cortex.2020.03.031
- Chalkia, A., Van Oudenhove, L., & Beckers, T. (2021). Corrigendum to "Preventing the return of fear in humans using reconsolidation update mechanisms: A verification report of Schiller et al. (2010)". *Cortex*, *139*, 309. http://doi.org/10.1016/j.cortex.2021.03.010
- Chambers, C. D. (2020). Verification Reports: A new article type at Cortex. *Cortex*, 129, A1–A3. https://doi.org/10.1016/j.cortex.2020.04.020
- Kredlow, M. A., Unger, L. D., & Otto, M. W. (2016). Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. *Psychological Bulletin*, *142*(3), 314–336. https://doi.org/10.1037/bul0000034
- McIntosh, R. D., & Chambers, C. D. (2020). The three R's of scientific integrity: Replicability, reproducibility, and robustness. *Cortex*, *129*, A4–A7. https://doi.org/10.1016/j.cortex.2020.04.019
- Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: Opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry*, *5*(12), 1032–1042. https://doi.org/10.1016/S2215-0366(18)30270-0
- Schiller, D., LeDoux, J. E., & Phelps, E. A. (2020). Reply to Beckers, McIntosh and Chambers on the verification of 'preventing the return of fear using retrieval-extinction in humans.' *PsyArXiv*. https://doi.org/10.31234/osf.io/jn6uw
- Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49–53. https://doi.org/10.1038/nature08637
- Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2018). Addendum: Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *562*(7727), E21. https://doi.org/10.1038/s41586-018-0405-7
- Schiller, D., Raio, C. M., & Phelps, E. A. (2012). Extinction training during the reconsolidation window prevents recovery of fear. *Journal of Visualized Experiments : JoVE*, (66), e3893. https://doi.org/10.3791/3893
- Zuccolo, P. F., & Hunziker, M. H. L. (2019). A review of boundary conditions and variables involved in the prevention of return of fear after post-retrieval extinction. *Behavioural Processes*, *162*, 39–54. https://doi.org/10.1016/j.beproc.2019.01.011