A meta-analysis of the effect of protein synthesis inhibitors on rodent fear conditioning

Clarissa F. D. Carneiro[1,2], Felippe E. Amorim[1] and Olavo B. Amaral[1]

[1] Institute of Medical Biochemistry Leopoldo de Meis, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

[2] Berlin Institute of Health (BIH) at Charité, BIH QUEST Center for Responsible Research, Berlin, Germany.

Abstract

Systematic reviews and meta-analyses have been increasingly recognized for their potential value in pre-clinical research, but their applications have not been extensively explored in behavioral neuroscience. In this work, we studied protein synthesis inhibition, a classic intervention used to disrupt fear learning, reconsolidation, and extinction in rodents, to explore how meta-analyses can identify potential moderators of its effect. We initially performed separate meta-analyses for different injection sites and target sessions to evaluate the effect of the intervention in various scenarios. Heterogeneity was further investigated by multilevel meta-regression models aggregating various sites, with article or research group as additional levels. We detected robust effects of protein synthesis inhibitors on training and reconsolidation, but not on extinction, possibly due to the lower number of studies on the latter. Our analyses identified some well-established moderators, such as intervention timing and reexposure duration. However, other factors proposed as boundary conditions for reconsolidation, such as memory age and training strength, were not associated with effect size. While our results point to the value of meta-analyses in consolidating findings from the literature, we believe that associations suggested by data synthesis should ideally be verified by well-powered, rigorous confirmatory experiments.

Introduction

Over the last decades, biomedical research has grown consistently; however, low reproducibility rates in some areas have shed doubt over its reliability[1–3]. Industry figures for the replicability of basic research have been as low as 11% and 21%[4,5]. In experimental neurology, only 30% of replications in spinal cord injury research were at least partly successful[6]. In basic cancer biology, a 6-year systematic effort reported that 46% of experiments from highly cited articles were reproduced, with effect sizes being 85% smaller in replications than in the original studies[2].

Systematic replication efforts, however, are cost- and labor-intensive endeavors and may not be feasible as a routine procedure in every area of science. Meta-analyses of the existing literature can thus be a useful and accessible tool to assess the reliability of effect estimates, investigate publication bias, explore sources of heterogeneity among studies and identify knowledge gaps in the literature[7]. This latter point is of particular interest to discovery science: as data synthesis is still rare in this field, conflicting results in the literature are usually not settled properly. Scientists tend to assume protocol discrepancies are responsible for divergences between studies, without considering the possibility that they might be due to random variation around statistical thresholds[8,9].

Methods for synthesizing data from basic biomedical research, however, are not as mature as in clinical research, where outcomes are typically fewer and more clearly defined. A single article in basic discovery science can include experiments exploring multiple questions, multiple methods addressing a single question, or multiple control

experiments testing boundary conditions (i.e. changing protocol parameters in order to establish the necessary conditions for an effect to be observed). This also leads to multiple experiments published in individual articles, as well as multiple articles by the same research groups, violating the assumption of independence between different results when synthesizing data.

While greater attention is present in other areas, meta-analysis of rodent behavioral studies is still incipient. Fear conditioning is a learning paradigm characterized by the association of a neutral conditioned stimulus (CS; usually a context or a tone) to an aversive unconditioned stimulus (US; for example, a footshock)[10]. Conditioning protocols vary in many ways, but the outcome used to measure memory is usually the same – the animal's freezing response to the CS in a test session. The paradigm has been widely used to dissect the anatomical and biochemical pathways involved in the acquisition and consolidation of fear memories. Furthermore, the same paradigm is employed to study reconsolidation, a process where a reactivated memories become labile and require protein synthesis to regain stability[11], and extinction, which involves the replacement of the original fearful association by a new, non-aversive one[12,13].

After a series of failed attempts to replicate findings on reconsolidation of fear conditioning in rodents[14,15], Schroyens *et al*. found evidence of publication bias in the field by reviewing the literature and contacting relevant authors[16]. In their sample, 80% of the published findings yielded statistically significant results, in contrast to only 20% of the unpublished findings shared with them[16]. The failed replications also raise the possibility that differences in protocols can determine the magnitude of the outcome. This suggests the need to determine the robustness of the effect of amnestic agents in different conditions, as well as to systematically investigate moderators that may explain the heterogeneity in results.

In this work, we use a classic intervention in mechanistic studies – the inhibition of protein synthesis to modulate fear learning, reconsolidation and extinction in mice and rats – as a model for exploring the possibilities of data synthesis in basic biomedical science. By selecting a well-established intervention with a large body of literature, we aim to investigate whether meta-analytic methods can detect consensual findings – such as the effect of protein synthesis inhibition on consolidation – and whether they can shed light on more controversial topics concerning its boundary conditions – such as the effect of reexposure duration, memory strength or memory age on reconsolidation. For this purpose, we explore different ways of aggregating studies within meta-analyses, as well as different approaches to deal with non-independence between experimental results.

Methods

The data collection and analysis protocol was posted prior to performing the work and is available online at https://osf.io/y8r2j. Details of the methods are presented below, highlighting additions to the prespecified protocol.

Study search and identification

PubMed, Web of Science and Scopus were searched for peer-reviewed publications using the terms "("protein synthesis" OR "anisomycin" OR "emetine" OR "cycloheximide" OR "protein synthesis inhibitor" OR "psi" OR "puromycin" OR "acetoxycycloheximide" OR "isocycloheximide" OR "chloramphenicol" OR "rapamycin") AND ("fear conditioning" OR "aversive conditioning" OR "conditioned fear" OR "fear memory" OR "fear learning" OR "aversive learning" OR "pavlovian conditioning" OR

"fear-motivated learning" OR "threat conditioning" OR "threat learning" OR "contextual conditioning" OR "auditory conditioning" OR "fear memory retrieval" OR "fear memory extinction" OR "fear memory reconsolidation" OR "freezing behavior")". The specific protein synthesis inhibitors included in the search were drawn from a pilot search using "protein synthesis" and "memory". In all searches, publication date was restricted to articles published until December 31st, 2018. The Scopus search was restricted to titles and abstracts, and the Web of Science search excluded book chapters, editorials, meeting abstracts and proceedings papers. Duplicates were excluded using Mendeley (version 1.17.10, Mendeley Ltd).

Study selection

The first screening step considered only titles and abstracts and excluded (i) articles not written in English, (ii) articles not presenting original results, such as reviews, and (iii) articles not describing fear conditioning experiments in mice or rats. Retracted articles were also excluded during this stage. This step was performed by two reviewers, and at least one of them had to include the reference for it to be taken to the next stage. If the title/abstract was not clear about the three criteria described above, the article was still included for further screening.

The second screening stage considered the full text of the article, including any supplementary material if available, and papers were included if they met the following criteria: (i) described the effects of a single dose of a protein synthesis inhibitor, administered either systemically or intracerebrally, (ii) administered protein synthesis inhibitors from 1 h before up to 24 h after a behavioral session (e.g. fear conditioning training or reexposure), (iii) assessed memory by a drug-free test using freezing behavior as a measure of fear, (iv) had a clearly defined control group to which the experimental group was compared, and (v) had available data on mean freezing and standard deviation (SD) or standard error of the mean (SEM) in the test session for each experimental group. In case multiple retrieval tests were performed, only the first drug-free test session was included in the meta-analysis. Drugs were considered as protein synthesis inhibitors only when directly affecting the process of translation itself. Therefore, drugs interfering with transcription or mRNA synthesis (such as actinomycin D) or affecting translation as a consequence of interference with an upstream signalling pathway (such as rapamycin) were excluded. Drugs interfering with the synthesis of specific proteins rather with the protein synthesis process as a whole, such as antisense oligonucleotides, were also excluded.

In the second screening stage, each article was evaluated by one of two investigators who indicated which experiments should be extracted for the meta-analysis. Extraction of the data was performed by the other reviewer, so that all included experiments were double-checked, and disagreements regarding inclusion were discussed with a third investigator until consensus was reached. After the completion of the screening phase, a random sample of 10% of excluded articles was double-checked by both investigators to obtain agreement levels for exclusions.

Data extraction

For each included article, we extracted data regarding article- and experiment-level features, as well as the quantitative results required for the meta-analyses. The

dataset used in all analyses is provided as **Supplementary File 1**. Additional data collected but not used in the analyses presented here are described in the original protocol and available at https://osf.io/agdte/.

For each article, we recorded the year of publication, impact factor (as per the respective year's Journal Citation Reports), number of citations (obtained at the end of data collection on June 25th 2020) and region of origin (defined by the corresponding author's affiliation). As articles were published within a range of over 20 years, citations were normalized as number of citations per year (number of citations/(2020 – year of publication)).

As risk of bias indicators for each article, we assessed if the following items were reported for any of the included experiments: randomization of animals to experimental groups, blinding or automation of outcome assessment (i.e. freezing measure), sample size calculation, statement regarding compliance with ethical regulations, and statement on conflict of interest.

The following experiment-level features were extracted: type of conditioning, training and testing protocols, reexposure protocols (when applicable), habituation and handling procedures, species, sex, age and number of animals per cage. Habituation and handling protocols were transformed to categorical dichotomous variables (i.e. reported or not). When housing information was presented as a range, we recorded the highest value. When age was presented as a range, we used the mean between minimum and maximum values.

Fear conditioning protocols were classified as "contextual" or "tone" according to the type of CS used in the training session. If contextual memory was evaluated by a test in the same context after tone conditioning, this was labelled as "contextual background". If presentation of the tone and footshock were separated by an interval, this was labelled as "tone-trace". We also recorded the number and intensity of footshocks presented at training and the time interval between sessions (training-testing, training-reexposure and reexposure-testing).

For each intervention, we extracted the active drug, dose and site of injection. These were later categorized as systemic, intra-amygdala, intra-hippocampal, intra-cerebroventricular or other (for other brain structures). We also recorded the time interval between the intervention and the behavioral session (training, reconsolidation or extinction) at which it was targeted.

Mean and SD or SEM for freezing levels and sample size were extracted for both the experimental and control groups. Numerical values were obtained from the text or legends when available, or directly from graphs when necessary using Gsys (version 2.4.6, Hokkaido University Nuclear Reaction Data Centre).

Missing Data

Corresponding authors were contacted for included articles in which data on the sample size of individual groups or the meaning of error bars (SD or SEM) were missing. Missing data on other protocol features were left blank.

Data exclusion

If authors did not respond to contact attempts, experiments with missing data on sample size were excluded from the analyses. When error bars were not defined as SD or SEM, data were included, but were assumed to represent SEM for a conservative estimate of variance. Additional tests of the same type (tone, tone-trace or contextual) on the same cohort of animals were excluded – i.e. meta-analyses only include the first drug-free test after intervention session (training or reexposure). However, if the same cohort was tested for tone and context memory (i.e. contextual background conditioning), both tests were included and sample size was divided between them for analysis. Similarly, when a group served as a control for more than one comparison, its sample size was divided by the number of extracted comparisons to avoid overrepresentation of this group in the meta-analyses.

During data extraction, we identified one article with duplicated data (same figure presented for different experiments) and one article in which mean freezing values in the text did not match the values in the figures. In both cases, the authors were contacted but could not retrieve the original data, so these experiments were excluded. After data extraction was completed, we also decided to exclude unclear or unusual test protocols, such as re-exposure protocols including the US or those in which the response to tone conditioning was not measured during the tone.

Analysis

All analyses were performed in R[17] (version 4.0.4). The complete code used is provided as **Supplementary File 2**.

Article Features

Frequency distributions of region of origin, impact factor, number of citations and year of publication are presented for the complete dataset. The percentage of risk of bias items reported was also calculated for the whole sample of articles.

Identification of research groups

Research groups were identified based on co-authorship networks of included articles[18]. Briefly, the list of authors was obtained for all articles included in our dataset to create a co-authorship graph using the Louvain method[19]. Modularity classes were then added to the dataset as an identification of research group. In case an article had authors attributed to different research groups, we assigned the article to the group to which most authors belonged.

Meta-Analyses

The random-effects model was chosen for all meta-analyses and fitted with the restricted maximum-likelihood (REML) estimator (except for multivariable meta-regressions, where the maximum-likelihood estimator was used due to computational limitations). For three-level meta-analyses, we opted for a fully random model, where random effects were used at both the article/group and experiment level. Main results present effect sizes as absolute mean differences in freezing, but we also performed sensitivity analyses using standardized mean differences (i.e. Hedges' *g*). All meta-analyses are based on sample sizes adjusted for the number of comparisons per control group.

As defined in the protocol, separate meta-analyses were performed for interventions in different sessions (training, reconsolidation and extinction) and sites (systemic, intra-amygdala, intra-hippocampus or intracerebroventricular). The distinction between reconsolidation and extinction in this case was based on the article's description of the experiment. We also included a meta-analysis for prefrontal cortex interventions on training, as these were identified in 3 articles (the minimum number specified in the protocol).

Also following the protocol, we performed a meta-analysis for memory reactivation, including all studies with interventions in reexposure sessions – i.e. both reconsolidation and extinction studies. Two experiments did not clearly state whether their protocol was meant to cause reconsolidation or extinction and were included in the reactivation analyses only. Of note, as most articles had more than one experiment included, they could be included in multiple meta-analyses, or contribute multiple experiments to a single meta-analysis.

Although this was not specified in the predefined protocol, we also performed aggregated meta-analyses, combining data from all injection sites for each intervention session and from all intervention sessions in each injection site. A meta-analysis of the complete dataset, including all injection sites identified and all intervention sessions was also included. This complete dataset, as well as the training and reactivation datasets, were used in moderator analyses due to its larger sample size.

For each meta-analysis, I^2 is presented as a measure of heterogeneity alongside the p-value for the Q-test for heterogeneity, which measures whether the variability in the observed effect sizes is larger than would be expected based on sampling variability alone. Low p values thus indicate that the observed differences in effect sizes between experiments or studies would be unlikely to occur by chance alone. Possible sources of variability were further explored through moderator analyses.

Moderator analyses

Univariable models were built for aggregated meta-analyses (i.e. complete, training and reactivation datasets) to investigate if individual methodological variables or study features were associated with effect size across experiments. Categorical variables used were conditioning type, reporting of habituation, reporting of handling procedures, species, sex, active drug, site of injection, randomization, blinding, sample size calculation, statement of compliance with regulatory requirements, presence of conflict of interest and region of origin. Continuous variables used were shock intensity, number of shocks, time between drug administration and training/reexposure, time between training and reexposure, reexposure duration (not applicable for interventions on initial conditioning), time between intervention session and test, mean age of animals, number of animals housed together, drug dose, impact factor and citations per year. Doses were z-scored by normalizing each value according to the mean value for the respective drug and route of administration (systemic or intra-cerebral) for each species (rat or mouse). Similarly, reexposure duration was z-scored according to conditioning type using the mean number of tones (for tone reexposure) or duration (for contextual reexposure) across studies to allow inclusion of both types of study in the same analysis.

For each univariable model, we present the R^2 value and the p-value of the Q test of moderators. In 3-level models, R^2 values were calculated as the difference in the sum of σ^2 values for both levels between the model with no moderators and the univariable model, divided by the sum of σ^2 values in the model with no moderators. In case this yielded a negative value, R^2 was considered to be zero.

As defined in the protocol, we restricted the number of moderators included in multivariable models to one variable for every 10 experiments. The following priority order was defined *a priori* in the protocol: re-exposure duration (not applicable to training meta-analyses), drug, dose, time between intervention session and test, time to reexposure (not applicable to training meta-analyses), species, shock intensity, number of shocks, sex and age of animals. Due to low prevalence of reporting, animal age was not included among the variables tested in any of the datasets. In addition to the predefined list, we added intervention session (applicable only to the complete dataset), injection site, conditioning type and time between drug administration and training/reexposure, which were selected on the basis of univariable models.

All combinations of variables from the selected list were tested in multivariable models, and the best models were ranked by corrected Akaike Information Criteria (AICc). For each best model selected (for complete, training and reactivation datasets with 2- and 3-level analyses), we decomposed the R^2 value for each moderator included. For this, we calculated the mean of the differences between R^2 from models with and without the moderator in all possible orders of moderator inclusion. Additionally, we performed a Q test of moderators for each variable (including all dummy variables for each categorical moderator) to obtain p-values for individual variables.

Publication Bias

Due to limitations in sample size, publication bias analyses were restricted to systemic, intra-amygdala and intra-hippocampal interventions on training and reconsolidation. This was performed using Egger's regression[20] and trim-and-fill analyses[21,22]. As recommended by the authors of the method[23], both L_0 and R_0 estimators were used to estimate the number of missing studies and to correct the effect estimate (and its p-value) after imputation of these studies.

As an additional exploratory measure of publication bias, excess significance tests[24] were also performed for each meta-analysis. These were based on the meta-analytic estimates using sample sizes as reported in the articles, with no adjustment for control group or test session redundancies.

Finally, as planned in the protocol, statistical power for a two-sample two-sided test was calculated for each experiment, considering its own sample size (without adjustments) and standard deviation, along with the respective meta-analytical effect size estimate expressed as absolute mean difference and an alpha of 0.05. We used this to perform correlation analyses between statistical power and effect size with article-level features. For this, we used the mean power and effect size of all experiments per article and calculated Spearman's correlation with impact factor, citations per year and risk of bias. In this case, risk of bias was used as a score calculated as 5 minus the sum of reported items.

Results

Study selection

Our search initially retrieved 934 articles, of which 91 met our inclusion criteria[11,13,25–113]. **Figure 1** presents a PRISMA flowchart for the inclusion and exclusion of articles[114].

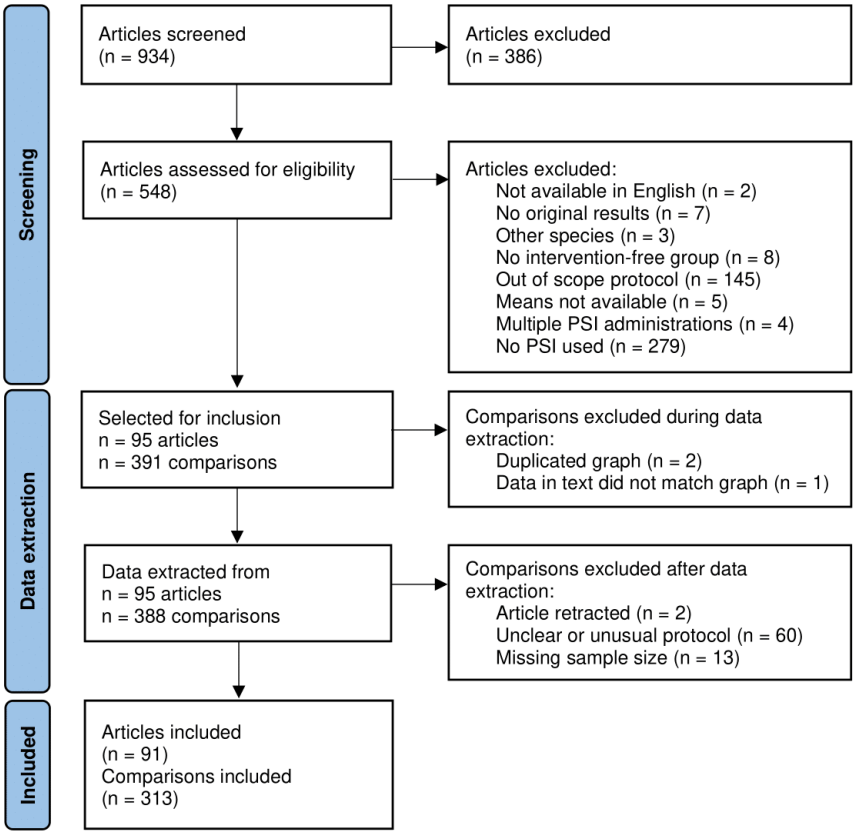

Figure 1 - Flowchart of screening stages. In the first stage of screening (n=934), two reviewers assessed titles and abstracts. If one reviewer included the article, it was selected for full-text screening. During this stage we did not record reasons for exclusion. The second screening stage (n=548) considered the full-text and supplementary material, with each being assessed by a single reviewer (no disagreements found in a double-screened sample of 36 articles). Additional exclusion of comparisons took place during or after data extraction (see Methods for details), leaving 313 comparisons from 91 articles to be included in the analyses. PSI: protein synthesis inhibitor.

Study description

Suppl. Figure 1 presents the frequency distribution of region of origin, year of publication, impact factor and citations for the full set of articles. Included articles were mostly from North America, published between 2005 and 2015 in journals with a median impact factor of 4.3. Frequency of reporting of risk of bias factors is presented in **Suppl. Table 1**. A fair number of articles reported blinded/automated assessment of outcomes, while randomization and sample size calculation were reported very rarely.

A summary of protocol variables is presented in **Table 1** (full data in **Suppl. Table 2)**. Most comparisons studied the inhibition of protein synthesis in training or reconsolidation sessions, while studies of extinction were less common. Systemic injections accounted for about a third of all comparisons, while intracerebral injections were mostly directed at the amygdala and hippocampus. Distribution between species

was balanced between rats and mice, with 94% of all comparisons including only male animals, and 86% of studies using anisomycin to inhibit protein synthesis.

Table 1 – Summary of protocol features of comparisons included in the analyses. For categorical variables, we show the number of comparisons per category, while for continuous variables we show the median and range (minimum and maximum) and the number of comparisons in which the variable was reported. Percentages refer to the total of 313 experiments, except for reexposure duration, where the total is 58 experiments using tone-cued re-exposure protocols and 87 experiments using contextual reexposure protocols. "Contextual background" refers to cases where conditioning was performed with tone cues, but testing/reexposure used the context as the CS. Complete data is presented in Suppl. Table 2.

Protocol variable	Value	N (%)
Intervention session	Training	168 (53.7%)
	Reconsolidation	119 (38.0%)
	Extinction	23 (7.3%)
	Reactivation	3 (1.0%)
Site of injection	Systemic	97 (31.0%)
	Amygdala	94 (30.0%)
	Hippocampus	71 (22.7%)
	Cerebral ventricles	9 (2.9%)
	Prefrontal cortex	9 (2.9%)
	Other	33 (10.5%)
Type of conditioning	Contextual	138 (44.1%)
	Tone	103 (32.9%)
	Contextual background	59 (18.8%)
	Tone-trace	13 (4.2%)
Number of footshocks	2 (1 – 76)	313 (100%)
Footshock intensity	0.75 mA (0.3 – 2)	310 (99.0%)
Reexposure duration	Context: 3 min (0.5 – 30)	84 (96.5%)
	Tone: 1 tone (1 – 15)	58 (100%)
Species	Rats	167 (53.4%)
	Mice	146 (46.6%)
Sex	Male	293 (93.6%)
	Both	18 (5.8%)
	Not informed	2 (0.6%)
Age	10 weeks (4 – 22)	122 (39.0%)
Animals per cage	1 (1 – 6)	242 (77.3%)
Active drug	Anisomycin	270 (86.3%)
	Cycloheximide	41 (13.1%)
	4EGI-1	2 (0.6%)

Meta-analyses

As planned *a priori*, we initially performed meta-analyses separated by intervention session (training, reconsolidation, extinction and reactivation) and site of drug administration (systemic, amygdala, hippocampus and cerebral ventricles). Most other intracerebral sites of drug administration were reported in less than 3 articles (**Suppl. Table 3**); thus, the only additional meta-analysis performed was for interventions on training with prefrontal cortex injections. In addition to these individual meta-analyses, we performed an exploratory meta-analysis with the complete dataset in order to explore predictors with greater power. **Table 2** summarizes these analyses, and forest plots are presented as **Suppl. Figures 2-4**.

Table 2 - Summary of meta-analyses. Effect sizes are expressed as absolute mean differences in freezing percentages between PSI-treated and control groups in the test session. Sample size is the

number of experiments. p values refer to main effect comparisons in random-effects meta-analyses (4th column) and Q-tests for heterogeneity (6th column). i.c.v.: intracerebroventricular, CI: confidence interval.

Dataset	Sample size	Effect size (% freezing) [95% CI]	Meta-analysis p-value	I^2	Q-test p-value
Training, systemic	57	-20.2 [-24.0, -16.3]	6.4×10^{-25}	82.8%	1.5×10^{-38}
Training, amygdala	38	-20.2 [-26.1, -14.3]	1.5×10^{-11}	66.6%	1.5×10^{-10}
Training, hippocampus	41	-14.1 [-18.9, -9.3]	9.0×10^{-9}	87.3%	1.9×10^{-23}
Training, i.c.v.	3	-22.0 [-53.5, 9.6]	0.173	80.9%	0.004
Training, prefrontal cortex	6	-10.7 [-20.5, -0.9]	0.033	0%	0.713
Reconsolidation, systemic	36	-18.6 [-23.1, -14.1]	4.2×10^{-16}	83.4%	4.4×10^{-35}
Reconsolidation, amygdala	47	-20.1 [-25.2, -15.0]	1.4×10^{-14}	78.8%	5.5×10^{-26}
Reconsolidation, hippocampus	20	-19.3 [-27.8, -10.9]	7.8×10^{-6}	89.6%	1.9×10^{-34}
Reconsolidation, i.c.v.	4	-12.5 [-43.9, 18.9]	0.435	88.5%	9.0×10^{-4}
Extinction, systemic	4	28.0 [21.4, 34.6]	1.2×10^{-16}	0%	0.771
Extinction, amygdala	8	-1.1 [-11.6, 9.4]	0.831	69.9%	0.001
Extinction, hippocampus	8	6.3 [-7.3, 19.9]	0.364	79.2%	0.002
Extinction, i.c.v.	2	21.0 [-12.2, 54.2]	0.214	84.8%	0.010
Reactivation, systemic	40	-14.2 [-20.1, -8.2]	3.3×10^{-6}	91.7%	1.3×10^{-66}
Reactivation, amygdala	56	-18.1 [-23.2, -13.0]	2.9×10^{-12}	82.4%	2.3×10^{-43}
Reactivation, hippocampus	30	-13.2 [-21.3, -5.2]	0.001	90.7%	1.9×10^{-59}
Reactivation, i.c.v.	6	0.9 [-21.7, 23.5]	0.939	91.6%	2.8×10^{-5}
Complete	313	-15.4 [-17.4, -13.3]	6.0×10^{-48}	87.0%	6.5×10^{-284}

We found significant negative effects on test freezing after protein synthesis inhibition in training and reconsolidation for almost all studied structures, with roughly comparable effect sizes. For extinction, however, an effect in the opposite direction (i.e. preventing freezing decline) was found only for systemic injections. That said, these results should be interpreted with caution, as there were far fewer studies on extinction than on training or reconsolidation. Since the reactivation dataset largely overlaps with the reconsolidation one, results were similar between them, but with smaller effect sizes for reactivation due to the inclusion of extinction experiments. Most analyses showed a large amount of heterogeneity between experiments. Results of the same analyses using standardized mean differences (Hedge's g) are shown as a robustness check in **Suppl. Table 4**.

Publication bias

To evaluate the presence of publication bias in our sample, we performed trim-and-fill analyses and Egger's regression for systemic, intra-amygdala and intra-hippocampal interventions on training and reconsolidation (other meta-analyses were not included due to their small sample size). **Figure 2** presents funnel plots with the results of Egger's regression and trim-and-fill analyses using the L_0 method (results with the R_0 method are presented in **Suppl. Table 5**). Unlike what was described by Schroyens et al. (2021) for various drugs in contextual fear reconsolidation, we did not find consistent evidence of publication bias for protein synthesis inhibitors: a significant correlation between effect size and precision was found only for intra-hippocampal interventions on reconsolidation, but in the opposite direction as would be expected from publication bias.

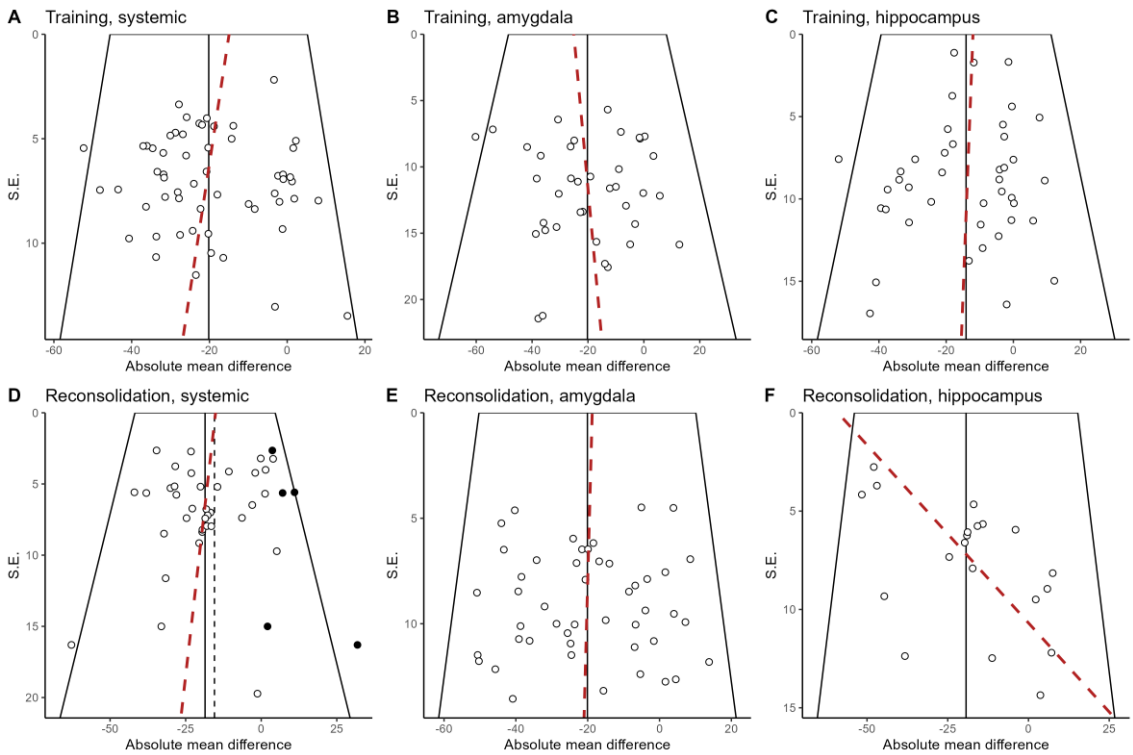

Figure 2 – Funnel plots and Egger's regression. Each observed study is represented by the white circles and the red dashed line represents Egger's regression line. Black circles represent missing studies imputed by the trim-and-fill L_0 method. **(A)** Training, systemic. Egger's regression p-value = 0.39. **(B)** Training, amygdala. Egger's regression p-value = 0.75. **(C)** Training, hippocampus. Egger's regression p-value = 0.39. **(D)** Reconsolidation, systemic. Egger's regression p-value = 0.24. Estimated number of missing studies = 5. **(E)** Reconsolidation, amygdala. Egger's regression p-value = 0.98. **(F)** Reconsolidation, hippocampus. Egger's regression p-value = 0.009. S.E.; Standard error.

Another potential source of bias in a meta-analysis is lack of statistical power in the original studies, which along with publication bias can lead to effect size inflation. In **Suppl. Figure 5,** we show the distribution of sample sizes and statistical power for training and reconsolidation experiments for systemic, intra-amygdala and intra-hippocampus interventions, calculated using the respective meta-analytic estimate as the reference effect size. Mean sample size (± SD) for all experiments combined was 9.1 ± 2.8 and mean statistical power for the corresponding meta-analytical effect size estimate was 0.66 ± 0.26. While differences in power for individual meta-analyses are heavily determined by differences between effect size estimates, experiments with

intracerebral injections also tended to have smaller sample sizes than those with systemic interventions. No correlation was found between mean effect size or statistical power and impact factor, citations or risk of bias score for articles (**Suppl. Figure 6**). Consistent with what we found by funnel-plot analysis, excess significance test did not detect evidence of publication bias (**Suppl. Table 6**).

Meta-regressions

A major challenge in performing meta-analyses of basic and pre-clinical research is defining the boundaries of a research question – i.e. how similar experiments should be for their data to be synthesized meaningfully. On one hand, including experiments with diverse protocols adds heterogeneity to the analyses, rendering their interpretation more difficult; on the other, this allows one to explore the impact of protocol variables on effect size.

A priori, we determined that our analyses would be divided by the target session of drug administration (training, reconsolidation or extinction) and by the site of drug administration (hippocampus, amygdala, i.c.v. or systemic), as presented in **Table 2** and **Suppl. Table 4**. However, to investigate whether aggregating data could provide additional information, we also combined all experiments for each intervention session (i.e. training, reconsolidation, extinction and reactivation), testing the injection site as a moderator (**Table 3, Suppl. Table 7**), and combined all experiments for each site of injection, testing the session as a moderator (**Table 3, Suppl. Table 8**). While the variance explained by injection site was relatively small (and driven mostly by structures other than the hippocampus and amygdala), the intervention session had a larger impact on effect sizes, particularly due to the influence of extinction experiments, in which the direction of effect was reversed. Nevertheless, a large amount of residual heterogeneity persisted even after adding these moderators.

Table 3 – Meta-regression models for testing intervention site and target session. Grey rows show the results of the models without moderators, while white lines show the results of meta-regressions. p-values refer to the Q-test for residual heterogeneity (3rd column) or the test of significance for the moderator (5th column). R^2 and p-values for multivariate meta-regression are presented in order, with the first row showing the results of intervention session and the second row showing those of injection site.

Dataset and moderators	I^2	p-value	R^2	p-value
Training, no moderators	81.9%	9.7×10^{-79}		
Training, injection site as moderator	80.3%	3.3×10^{-69}	5.2%	0.058
Reconsolidation, no moderators	85.9%	2.3×10^{-123}		
Reconsolidation, injection site as moderator	83.2%	1.2×10^{-92}	17.6%	2.3×10^{-4}
Extinction, no moderators	84.4%	5.0×10^{-12}		
Extinction, injection site as moderator	70.0%	1.3×10^{-5}	37.1%	0.018
Systemic, no moderators	88.2%	3.7×10^{-104}		
Systemic, intervention session as moderator	82.4%	4.2×10^{-70}	36.6%	3.8×10^{-10}
Amygdala, no moderators	77.8%	9.2×10^{-51}		
Amygdala, intervention session as moderator	74.1%	6.4×10^{-36}	16.5%	0.002
Hippocampus, no moderators	90.5%	1.0×10^{-89}		
Hippocampus, intervention session as moderator	88.3%	1.6×10^{-55}	17.0%	0.004
i.c.v., no moderators	91.7%	1.8×10^{-9}		
i.c.v., intervention session as moderator	84.9%	6.6×10^{-6}	14.1%	0.198
Other, no moderators	70.0%	2.0×10^{-12}		
Other, intervention session as moderator	57.0%	8.1×10^{-6}	41.1%	4.9×10^{-4}
Complete, no moderators	86.9%	2.9×10^{-282}		
	81.9%	2.4×10^{-195}	20.9%	7.6×10^{-14}

Complete, intervention session and injection site as moderators			7.4%	9.0×10^{-5}

A final possibility to increase sample size for meta-regression is by combining both intervention sessions and sites of drug administration using the complete dataset. To assess the impact of this merging, we used a multivariable meta-regression model considering both intervention session and site of injection (**Table 3, Suppl. Table 9**). Although both moderators account for a relevant amount of heterogeneity, we considered it reasonable to use the complete dataset to explore additional protocol variables in meta-regression, due to the greater sample size achieved. Thus, from here onward, we restrict our analyses to the complete dataset and to datasets with all interventions (irrespectively of site) on training and on reactivation (including reconsolidation, extinction and reexposure interventions).

Multilevel meta-analyses

The standard random-effects models in the meta-analyses presented above consider two sources of variability: the real variance among different experiments and the sampling error of each experiment. Both of these, however, are nested within at least two additional sources of variability: that between articles from which experiments are obtained and that between research groups publishing the articles. As multiple experiments within a paper and/or by a research group are not completely independent of each other, not taking these levels into account may bias the analyses towards results of particular articles or research groups that are overrepresented in the sample.

To test the importance of including these additional levels of variability, we applied an authorship network-based method[18] to identify research groups in the complete set of articles included (see Methods for details). Forty-two groups were identified by modularity analyses, as shown in **Suppl. Fig. 7**. The distribution of experiments within articles and research groups can be visualized in **Suppl. Fig. 8**. Although most articles and research groups contribute only a few experiments to the sample, a few are distinctly overrepresented in the sample (with a single research group contributing 58 experiments).

We then built additional models assuming random effects both within and between either articles or research groups (**Table 4**). This led to minor changes in estimates, and variance was larger between experiments than between either articles or research groups for all models. Nevertheless, in the reactivation dataset, these additional levels accounted for a reasonable fraction of the observed heterogeneity. The amount of variance explained by articles and research groups can also be observed by including articles and research groups as moderators (i.e. fixed factors) in meta-regressions (**Suppl. Table 10**), although this does not allow correction of effect estimates, as occurs when they are included as levels (i.e. random factors).

We also built a four-level model with articles and research groups as different levels (**Suppl. Table 11**). However, variance estimations were much less precise in this case, indicating a lack of clear distinction between each level's contribution to total heterogeneity.

Table 4 – Three-level meta-analyses. Two 3-level models were built, considering nesting within clusters that represented either articles or research groups, which are modeled as additional random factors. The two-level (i.e. standard random-effects) model is also included

for comparison. Effect sizes are in absolute mean differences. CI, confidence interval; Exp., experiments; cl., clusters.

Model	Sample size		Effect size (% freezing) [95% CI]	p-value	Variance [95% CI]		I^2	
	Exp.	cl			Between experiments	Between clusters	Exp.	Cluster
Complete dataset								
2-level	313	0	-15.35 [-17.4, -13.3]	7.9×10^{-48}	265.8 [215.7, 325.2]	-	85.8%	-
3-level (article)	313	91	-15.14 [-17.9, -12.4]	8.2×10^{-28}	202.3 [157.0, 260.5]	68.3 [25.2, 135.1]	65.2%	22.0%
3-level (group)	313	42	-15.27 [-18.6, -12.0]	1.2×10^{-19}	237.2 [188.9, 296.4]	40.4 [3.7, 125.8]	74.7%	12.7%
Training dataset								
2-level	168	0	-17.35 [-19.8, -14.9]	2.9×10^{-44}	174.6 [122.8, 230.2]	-	81.9%	-
3-level (article)	168	57	-17.19 [-19.8, -14.5]	4.5×10^{-37}	163.3 [114.8, 228.9]	11.4 [0, 58.1]	76.6%	5.4%
3-level (group)	168	29	-17.31 [-19.8, -14.9]	4.4×10^{-44}	174.6 [128.5, 236.9]	1.1×10^{-6} [0, >10.0]	81.9%	0%
Reactivation dataset								
2-level	145	0	-13.24 [-16.7, -9.8]	3.1×10^{-14}	361.6 [280.8, 497.8]	-	89.5%	-
3-level (article)	145	56	-13.06 [-17.5, -8.6]	6.8×10^{-9}	249.1 [171.2, 363.6]	119.8 [32.2, 265.5]	60.6%	29.1%
3-level (group)	145	28	-12.21 [-17.9, -6.5]	2.5×10^{-5}	296.9 [215.4, 413.6]	96.9 [6.6, 309.2]	68.0%	22.2%

Univariable meta-regressions

To explore the impact that individual features in the experiments might have on effect sizes, we initially performed univariable meta-regressions using the two-level model for all protocol and article variables collected, including risk of bias measures. Results are summarized in **Figure 3A**, with a more extensive description available in **Suppl. Tables 12-14**.

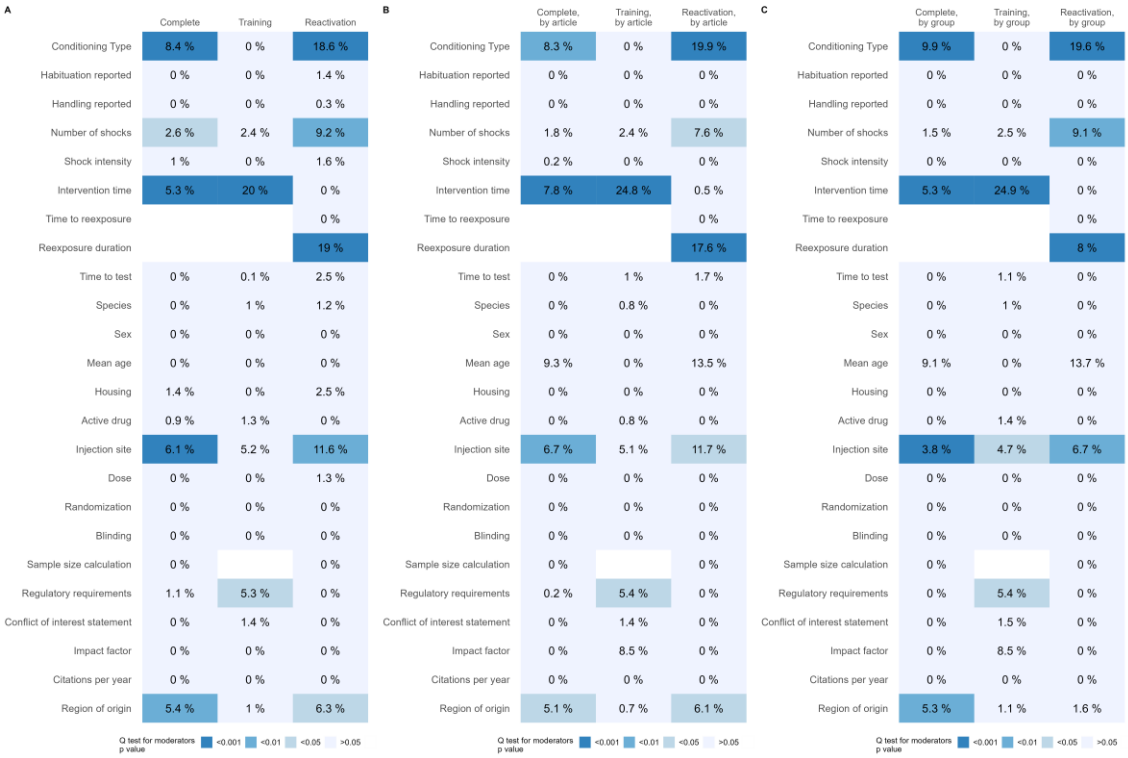

Figure 3 – Summary of univariable meta-regression models. Each column represents a dataset, while lines correspond to each variable. Cell colors represent p-values for Q tests of the respective variable as a moderator and numbers within cells are the R^2 value for the moderator in a univariable model. White cells indicate that the variable was not applicable or unfeasible to include in the model due to lack of variability. Reexposure duration is based on z-scored values for reexposure to tone (measured in number of tones) and to context (measured in minutes), based on the mean and standard deviation for each type of conditioning. Doses were also z-scored based on the mean and standard deviation of each drug and route of administration for each species. **(A)** Two-level models. **(B)** Three-level models, accounting for nesting of experiments within articles. **(C)** Three-level models, accounting for nesting of experiments within research groups. Full details for each model, including betas and directionality of effects, are shown on **Suppl. Tables S12-S21**.

For interventions in the training session, the time between drug administration and training (i.e. intervention time) was found to be a significant moderator of effect size. As can be observed in **Suppl. Figure 9**, this relationship is largely driven by injections performed 24 hours after the training session, which yield near-null effects.

In the reactivation sample, reexposure duration and number of tones (which were z-scored and included as a single variable) predictably influenced effect size. As there was little variation in the number of tones across experiments, the moderator effect can be attributed mainly to contextual fear conditioning experiments, in which effects on freezing were negative with under 10 minutes of reexposure (suggesting interference with reconsolidation), variable between 10 and 20 minutes, and mostly positive with 30 minutes of reexposure (suggesting effects on extinction) (**Suppl. Figure 10A-B**). This is confirmed by performing separate meta-regressions based on the duration of reexposure to context or number of tones (**Suppl. Table 15**), which show that the effect of reexposure duration on effect sizes is much more robust for contextual conditioning than for tone conditioning, in which experiments exploring larger number of tones were less common.

Additionally, conditioning type by itself also impacted effect sizes, with tone conditioning leading to larger effects on reconsolidation. As can be observed in **Suppl.**

Figure 10C, the impact of conditioning type is probably due to the large number of near-null effects for contextual fear reconsolidation, as well as to a greater number of extinction experiments using contextual conditioning. A higher number of shocks was also associated with larger effects on reconsolidation. That said, as can be seen from **Suppl. Figure 11A**, this association is mostly driven by a few experiments with 60 shocks and large effects, while an opposite trend is observed for experiments with up to 10 shocks.

Other proposed boundary conditions for reconsolidation, such as time between training and reexposure, were not detected as moderators in the model. That said, **Suppl. Figure 11B** shows that, for the few experiments with more than 20 days between training and reexposure, effects seem to be consistently smaller than those with shorter intervals, where there is greater variability in results. It is also worth noting that extinction experiments generally used short intervals, which might have obscured an effect of this variable on reconsolidation. Surprisingly, drug dose was also not found to be a significant moderator, although this could be due to the relatively narrow range of doses used in most studies when specific drugs, administration routes and species were taken into account (**Suppl. Figure 12**).

Next, we performed the same meta-regressions for three-level models, accounting for the nesting of experiments within articles (**Figure 3B** and **Suppl. Tables 16-18**) and within research groups (**Figure 3C** and **Suppl. Tables 19-21**). In these models, the amount of heterogeneity accounted for most moderators is lower, suggesting that some of the moderator effects might be explained by the article/research group of origin instead. However, the main findings persist – as in the case of injection time in training, and of conditioning type and reexposure duration in reactivation.

Multivariable meta-regressions

Lastly, we explored multivariable models to investigate whether the moderator effects detected in univariable regressions could be due to confounding by other variables. We started with the complete dataset and the original variables that separated the analyses (injection site and intervention session), the two variables with the largest effects in univariable models (conditioning type and intervention time) and 7 other variables selected *a priori* (active drug, dose, time to test, species, number of shocks, shock intensity and sex). Missing data for some of these variables led to the exclusion of 36 experiments, leaving 277 to be used in these analyses.

In the training dataset, including the same variables listed above (except for intervention session) led to the exclusion of 11 experiments, leaving 157 available for the analyses. In the reactivation dataset, the predetermined list of variables included two additional variables: reexposure duration and time between test and reexposure. This led to exclusion of 19 experiments, leaving 126 available.

After testing all possible variable combinations in the 2-level and 3-level models (2,048 in the complete dataset, 1,024 in the training and reactivation datasets), we ranked the resulting models by the Akaike Information Criteria (AICc). For each dataset and clustering option, we present a summary of the best models in **Table 7** and effect sizes are presented in **Suppl. Tables 22-30**.

Table 7 – Best multivariable meta-regression models for each dataset and clustering option. Only the model with lowest AICc is presented for each scenario. Weight refers to the Akaike weights. Q-test p-

values refer to the test for all included moderators. Effect sizes and p values for individual moderators in each model are presented in **Suppl. Tables 22-30.**

Dataset	Clusters	Included variables	AICc	Weight	I^2	R^2	Q-test p-value
Complete	2-level	Intervention session Site of injection Conditioning type Active drug Number of shocks Intervention time	2328.1	0.060	76.4%	45.1%	9.6×10^{-24}
Complete	3-level (article)	Intervention session Site of injection Active drug Number of shocks Intervention time	2328.8	0.056	79.2%	44.5%	4.6×10^{-22}
Complete	3-level (group)	Intervention session Site of injection Active drug Number of shocks Intervention time	2326.7	0.081	79.7%	44.0%	2.1×10^{-22}
Training	2-level	Site of injection Intervention time	1299.2	0.037	73.1%	27.7%	1.9×10^{-6}
Training	3-level (article)	Site of injection Intervention time	1301.4	0.036	75.4%	30.7%	1.9×10^{-6}
Training	3-level (group)	Site of injection Intervention time	1301.4	0.037	75.4%	30.1%	1.2×10^{-6}
Reactivation	2-level	Site of injection Reexposure duration Number of shocks	1090.9	0.117	81.1%	45.2%	5.1×10^{-14}
Reactivation	3-level (article)	Site of injection Reexposure duration Number of shocks	1092.4	0.124	82.6%	43.6%	7.3×10^{-13}
Reactivation	3-level (group)	Site of injection Reexposure duration Number of shocks	1093.2	0.122	82.3%	47.4%	5.1×10^{-14}

Notably, multivariable models did not include some article-level variables detected as significant by univariable models, such as presence of a regulatory requirements statement and region of origin, suggesting that these effects were likely due to confounding. Nevertheless, most variables related to the experimental protocol itself were preserved. These results can also be visualized by the importance of all variables in each set of models (**Suppl. Figure 13**).

For each of these models we decomposed the R^2 value for each of the variables selected to investigate the contribution of each one (**Figure 4**). Once more, it is notable that R^2 values in the reactivation set are lower for most variables in multilevel models, suggesting that some of the correlation between these variables and effect sizes is driven by confounding from the use of specific reexposure protocols by particular articles or groups. Conditioning type in particular was not detected as a moderator when additional levels are considered, suggesting that this effect might have been driven by research groups who preferentially use tone or context conditioning. This does not seem to be the case in the training set, where including articles or groups as levels actually increased the effect of intervention time and site of injection.

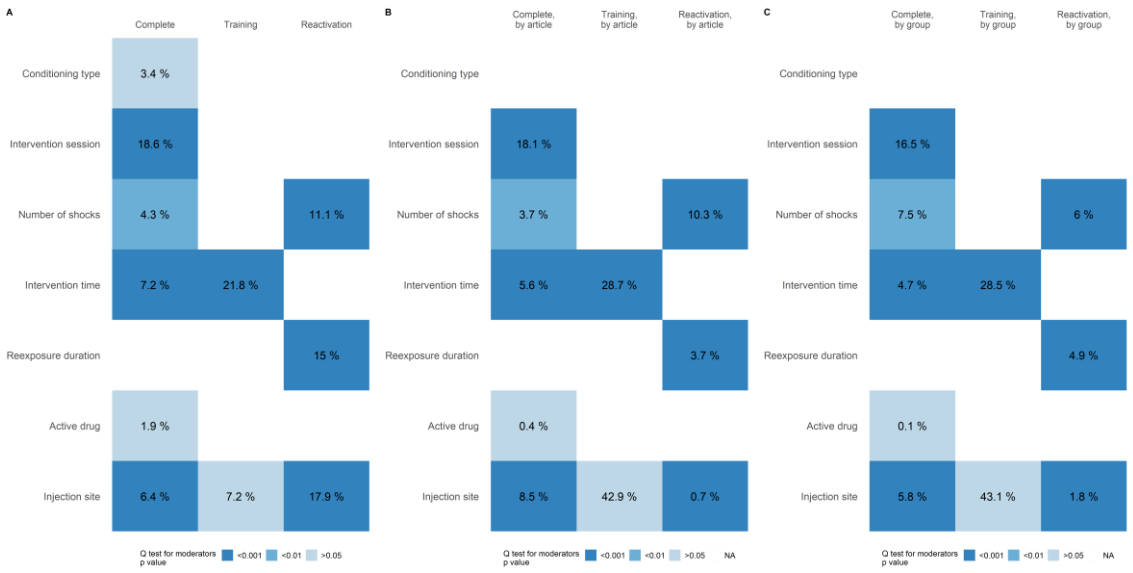

Figure 4 – Summary of multivariable meta-regression models. Each column represents a dataset and lines are variables selected in the best model for each scenario. Cell colors represent the p value for the Q test for each moderator in the meta-regression and numbers within are the decomposed R^2 value for the moderator (see Methods). Reexposure duration is based on z-scored values for reexposure to tone (measured in numbers) and to context (measured in minutes), based on the mean and standard deviation for each type of conditioning. **(A)** Two-level models. **(B)** Three-level models, accounting for nesting of experiments within articles. **(C)** Three-level models, accounting for nesting of experiments within research groups.

Discussion

In this study, we explored different meta-analysis methods to analyze the effect of protein synthesis inhibitors on fear conditioning. Beyond the results, this effort also leads to some insights on how to apply existing meta-analysis methods to discovery science, which presents challenges to data synthesis.

Effect of protein synthesis inhibitors on consolidation, reconsolidation and extinction

One particular focus was to investigate different ways to aggregate studies. Combining them by intervention session and by injection site showed a very robust effect for interventions in the three most studied sites (systemic injections, hippocampus and amygdala) on training and reconsolidation. This is unsurprising, as the requirement of protein synthesis for consolidation has been well accepted for decades[115]. More recently, protein synthesis inhibitors also became the paradigmatic intervention to demonstrate the existence of reconsolidation[11,116].

On the other hand, the effect of protein synthesis inhibition on fear extinction was not as robust. Findings on the role of protein synthesis in extinction, in fear conditioning and beyond[117,118], were not consensual from the start[112]. Moreover, when compared to reconsolidation, much less evidence on the topic is available. Thus, our meta-analysis found a significant effect only for systemic interventions. That said, most null effects on extinction from amygdala interventions come from a single study[90], whereas the hippocampus analysis is heavily influenced by a study which, although investigating extinction according to the authors, would be considered a reconsolidation experiment by most standards[106]. Our meta-analysis thus suggests that it would be worthwhile to conduct further investigations into the role of protein synthesis in extinction in both brain structures.

Even though the training and reconsolidation data are robust on aggregate, there is great heterogeneity among individual experiments for virtually all meta-analyses performed. When articles were included as an additional level, most of this heterogeneity was observed between different experiments within studies, perhaps because behavioral neuroscience articles typically include experiments as tests of boundary conditions or negative controls. This presents a challenge for meta-analysis, as selection criteria will commonly lead to inclusion of many of these experiments, which are likely to bias effect size estimates downwards.

Aggregating studies from all injection sites or from interventions in training and reactivation did not cause much of an increase in heterogeneity, perhaps because it was already large to start with. As the increased sample sizes obtained provided greater power for identifying moderators, we used this approach for our meta-regressions, in a decision that was not prespecified in the original protocol.

Moderators and boundary conditions

Some of the included moderators provide sanity checks of whether meta-regression captures consensual knowledge. It is widely accepted, for instance, that the effect of protein synthesis inhibitors is time-dependent, occurring when they are administered up to a few hours after learning[110,119]. Accordingly, in the training dataset, we found the time between training and drug administration to be a significant moderator in all models tested.

Similarly, it is well accepted that the occurrence of reconsolidation or extinction depends on the degree of mismatch between the training session and the reactivation, which is classically modulated by varying the number of CS presentations[120,121] or the duration of reexposure[13,122]. Our measure of reexposure duration, which incorporated both of these manipulations, was indeed found to be a significant moderator in the reactivation dataset. Interestingly, however, it had a much smaller effect size when research group was included as a level in the analysis.

Other postulated boundary conditions for reconsolidation, however, were not observed as clearly. Many studies suggest that training strength can determine the occurrence of reconsolidation, extinction or neither[13,123,124]. This relation, however, is not linear: while very low training strength favors extinction rather than reconsolidation[123], very strong memories have been suggested to be more resistant to destabilization[124]. Contrary to that, we found that a higher number of shocks led to a greater effect of protein synthesis inhibitors; that said, this effect was largely driven by a few experiments with a very large number of shocks. In contrast, shock intensity did not appear to be a relevant moderator, contradicting studies where administration of glucocorticoids after reactivation impaired retrieval only when stronger shocks were used[125].

Another proposed boundary condition for reconsolidation is memory age[126], but this has not been consistent across studies[49] and may depend on reexposure conditions[13]. Our meta-analysis did not find an influence of memory age on the effect of protein synthesis inhibitors on reactivation; This runs counter to widely cited studies showing that older memories are resistant both to protein synthesis inhibitors[13,85,126] and other drugs[127]. Nevertheless, it is worth pointing out that our model does not include interactions between moderators; thus, an effect could have been obscured by confounding. It is also possible that this effect is limited to very remote memories:

studies with memories over 20 days-old had smaller effects in our sample, but this was offset by large effects for those between 1- and 20 days-old.

Type of conditioning (tone vs. contextual) was consistently detected as a significant moderator for reconsolidation effects in our univariable models, mostly driven by frequent null results in contextual conditioning experiments. Importantly, however, the effect we found was nearly absent from multivariable models, perhaps due to its interaction with site of injection, as both types of conditioning tend to be studied with interventions in different brain structures.

Although one would expect that drug dose would be a relevant moderator for a pharmacological intervention, this was not the case in our meta-analysis, probably due to the low variability in the doses used for each drug, species and route of administration (which were used as normalizers in the analysis). This is likely due to the fact that, as protein synthesis inhibitors have been in study for many years, effective doses are well established and rarely varied in more recent studies.

The changes in the relative importance of different moderators when multivariable and multilevel models are used underscore the importance of these methods to minimize confounding. Interestingly, the effect of intervention time, perhaps our best-established moderator, became stronger when multiple levels were added. Conversely, seemingly spurious factors such as region of origin or presence of a conflict-of-interest statement disappeared when research group was used as a level.

Heterogeneity and replication failures

Even after moderators and additional levels were included, however, much heterogeneity remained, suggesting that classic boundary conditions are not sufficient to explain discrepancies in the literature. This echoes recent failures in replicating post-retrieval amnesia studies, despite the use of similar protocols[14,128]. Minor differences in reactivation protocols, animal traits[129], housing conditions[130] or interactions between moderators could explain this additional heterogeneity. Nevertheless, it is also possible that heterogeneity and replication failures are due to biases in the literature.

In this sense, the prevalence of randomization and sample size calculations in our sample was low, in accordance with previous surveys of preclinical studies[131]. Encouragingly, however, we found a higher level of blinding than reported in other areas[131], mostly due to the use of automated freezing detection. Another interesting finding is the near absence of studies using female animals, which has also been pointed out in previous studies of fear conditioning[132] and deviates markedly from current recommendations[133,134].

Another explanation for divergences in the literature is random error, which can lead to effect size vibration across statistical thresholds, particularly when power is low. We found the average power of individual studies to detect the respective meta-analytical effect size estimate to be 63%, which is in line with a previous estimate for fear conditioning[132]. This means that around one third of studies should fail to find significant differences between groups even when a real effect exists, suggesting that lack of replication based on statistical significance criteria is expected to occur frequently.

Interestingly, we found little evidence of publication bias, in contrast to what was found by Schroyens et al (2021) for anisomycin, propranolol, MK-801 and midazolam in contextual fear reconsolidation. The evidence for publication bias in their study is

weaker for anisomycin, however, suggesting that it may be a smaller problem in the literature on protein synthesis inhibitors. Nevertheless, our ability to detect publication bias may have been limited by experiments testing boundary conditions or negative controls, as bias may act on the opposite direction in these cases (i.e. suppressing statistically significant results). Publication bias assessments can also be limited by true heterogeneity[23], which is quite large in our sample.

Conclusion and future remarks

In conclusion, we have found that meta-analysis and meta-regression can capture some well-established findings in the fear conditioning literature. This includes a robust effect of protein synthesis inhibitors on consolidation and reconsolidation and its dependence on moderators such as timing of post-training interventions and the degree of reexposure in reactivation experiments. Although our analysis only includes studies published up to 2018, these main effects are robust and would be unlikely to change much if more recent studies were included.

Nevertheless, we also found some knowledge gaps, particularly on the effects of protein synthesis inhibition in specific brain structures on extinction. Moreover, not every boundary condition commonly mentioned in the literature was confirmed, including factors such as number of shocks, shock intensity and memory age. This may indicate that the evidence for some of these is more controversial than suggested in the literature, or alternatively that meta-regression might fail to capture real effects due to noise, non-linearity or interactions between variables (which were not assessed in our analyses). This leads to a conundrum – when meta-regression doesn't agree with individual studies that test a particular moderator, which of them should we trust?

Unfortunately, there is no easy answer. On one hand, individual studies are typically underpowered and subject to many biases. On the other, meta-regression findings are observational and inevitably prone to confounding. The use of multilevel models and multivariable regression can mitigate this problem but will not fully eliminate it. Moderators found by meta-regression should thus be considered tentative and ideally tested in empirical studies designed for this purpose. Although data synthesis can be useful for evaluating the primary literature, it incorporates its biases, and cannot fully substitute for confirmatory, high-powered replication studies[135] – which should be regarded as the preferred option for solving controversies between meta-analyses and primary studies.

A final advantage of empirically testing findings from meta-analysis is to compare the predictions of different types of models in terms of their accuracy. Although we argue in favor of using multilevel and multivariable models, these can be unfeasible or inaccurate when the number of studies is low. It is thus important to evaluate how much is lost by traditional two-level models, which in our work were in general agreement with multilevel ones.

All of these issues will only be resolved when systematic iteration between data synthesis and experimentation becomes more common in discovery science. This requires engaging researchers in systematic reviews, as well as making these more efficient through automated tools[136] and better structuring of primary data. It also involves building up infrastructure for confirmatory research that can validate meta-analytical findings empirically[135]. Ultimately, such interplay may produce a virtuous cycle that can improve the robustness of basic science beyond what can be achieved with the current model of isolated experiments and narrative synthesis.

Acknowledgments

We'd like to thank Daniel Menezes for assistance in the early stages of screening, Thiago Moulin for assistance with co-authorship graphs, Kleber Neves for general assistance with analyses in R and Jonathan Lee for opinions on the manuscript.

Furthermore we thank the recommender and reviewers from Peer Community In Neuroscience for their assessments and suggestions to improve this manuscript.

References

1. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*. 2015;349(6251):aac4716. doi:10.1126/science.aac4716

2. Errington TM, Mathur M, Soderberg CK, et al. Investigating the replicability of preclinical cancer biology. *eLife*. 2021;10. doi:10.7554/eLife.71601

3. Amaral OB, Neves K, Wasilewska-Sampaio AP, Carneiro CF. The brazilian reproducibility initiative. *eLife*. 2019;8. doi:10.7554/eLife.41602

4. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012;483(7391):531-533. doi:10.1038/483531a

5. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10(9):712. doi:10.1038/nrd3439-c1

6. Steward O, Popovich PG, Dietrich WD, Kleitman N. Replication and reproducibility in spinal cord injury research. *Exp Neurol*. 2012;233(2):597-605. doi:10.1016/j.expneurol.2011.06.017

7. Vesterinen HM, Sena ES, Egan KJ, et al. Meta-analysis of data from animal studies: A practical guide. *J Neurosci Methods*. 2014;221:92-102. doi:10.1016/j.jneumeth.2013.09.010

8. França TF, Monserrat JM. Reproducibility crisis in science or unrealistic expectations? *EMBO Rep*. 2018;19(6):e46008. doi:10.15252/embr.201846008

9. Stanley DJ, Spence JR. Expectations for Replications: Are Yours Realistic? *Perspect Psychol Sci J Assoc Psychol Sci*. 2014;9(3):305-318. doi:10.1177/1745691614528518

10. Wotjak CT. Sound check, stage design and screen plot – how to increase the comparability of fear conditioning and fear extinction experiments. *Psychopharmacology (Berl)*. 2019;236(1):33-48. doi:10.1007/s00213-018-5111-5

11. Nader K, Schafe GE, Le Doux JE. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*. 2000;406(6797):722-726. doi:10.1038/35021052

12. Bouton ME. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol Psychiatry*. 2002;52(10):976-986. doi:10.1016/S0006-3223(02)01546-9

13. Suzuki A, Josselyn SA, Frankland PW, Masushige S, Silva AJ, Kida S. Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *J Neurosci*. 2004;24(20):4787-4795. doi:10.1523/JNEUROSCI.5491-03.2004

14. Luyten L, Schnell AE, Schroyens N, Beckers T. Lack of drug-induced post-retrieval amnesia for auditory fear memories in rats. *BMC Biol*. 2021;19(1):17. doi:10.1186/s12915-021-00957-x

15. Schroyens N, Alfei JM, Schnell AE, Luyten L, Beckers T. Limited replicability of drug-induced amnesia after contextual fear memory retrieval in rats. *Neurobiol Learn Mem*. 2019;166:107105. doi:10.1016/j.nlm.2019.107105

16. Schroyens N, Sigwald EL, Van Den Noortgate W, Beckers T, Luyten L. Reactivation-dependent amnesia for contextual fear memories: Evidence for publication bias. *eNeuro*. 2021;8(1):1-16. doi:10.1523/ENEURO.0108-20.2020

17. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2022. https://www.r-project.org/

18. Moulin TC, Amaral OB. Using collaboration networks to identify authorship dependence in meta-analysis results. *Res Synth Methods*. 2020;11(5):jrsm.1430. doi:10.1002/jrsm.1430

19. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp*. 2008;2008(10):P10008. doi:10.1088/1742-5468/2008/10/P10008

20. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J*. 1997;315(7109):629-634. doi:10.1136/bmj.315.7109.629

21. Duval S, Tweedie R. Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*. 2000;56(2):455-463. doi:10.1111/j.0006-341X.2000.00455.x

22. Duval S, Tweedie R. A Nonparametric “Trim and Fill” Method of Accounting for Publication Bias in Meta-Analysis. *J Am Stat Assoc*. 2000;95(449):89-98. doi:10.1080/01621459.2000.10473905

23. Duval S. The Trim and Fill Method. In: Rothstein H, Sutton A, Borenstein M, eds. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons Ltd; 2005.

24. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials*. 2007;4(3):245-253. doi:10.1177/1740774507079441

25. Frankland PW, Josselyn SA, Anagnostaras SG, Kogan JH, Takahashi E, Silva AJ. Consolidation of CS and US representations in associative fear conditioning. *Hippocampus*. 2004;14(5):557-569. doi:10.1002/hipo.10208

26. Runyan JD, Dash PK. Inhibition of hippocampal protein synthesis following recall disrupts expression of episodic-like memory in trace conditioning. *Hippocampus*. 2005;15(3):333-339. doi:10.1002/hipo.20055

826 27. Motanis H, Maroun M. Differential involvement of protein synthesis and actin
827 rearrangement in the reacquisition of contextual fear conditioning. *Hippocampus*.
828 2012;22(3):494-500. doi:10.1002/hipo.20915

829 28. XinChun J, XueLian Q, XiaoFei Y, BaoMing L. Protein synthesis inhibition in the basolateral
830 nucleus of amygdala facilitates extinction of auditory fear memory. *Chin Sci Bull*.
831 2007;52(18):2532-2542. doi:10.1007/s11434-007-0393-8

832 29. Wanisch K, Tang J, Mederer A, Wotjak CT. Trace fear conditioning depends on NMDA
833 receptor activation and protein synthesis within the dorsal hippocampus of mice. *Behav*
834 *Brain Res*. 2005;157(1):63-69. doi:10.1016/j.bbr.2004.06.009

835 30. Yim AJ, Moraes CRG, Ferreira TL, Oliveira MGM. Protein synthesis inhibition in the
836 basolateral amygdala following retrieval does not impair expression of morphine-
837 associated conditioned place preference. *Behav Brain Res*. 2006;171(1):162-169.
838 doi:10.1016/j.bbr.2006.03.031

839 31. Patricio Casanova J, Madrid C, Contreras M, et al. A role for the interoceptive insular
840 cortex in the consolidation of learned fear. *Behav BRAIN Res*. 2016;296:70-77.
841 doi:10.1016/j.bbr.2015.08.032

842 32. Abate G, Colazingari S, Accoto A, Conversi D, Bevilacqua A. Dendritic spine density and
843 EphrinB2 levels of hippocampal and anterior cingulate cortex neurons increase
844 sequentially during formation of recent and remote fear memory in the mouse. *Behav*
845 *Brain Res*. 2018;344:120-131. doi:10.1016/j.bbr.2018.02.011

846 33. Rizzo V, Touzani K, Raveendra BL, et al. Encoding of Contextual Fear Memory Requires De
847 Novo Proteins in the Prelimbic Cortex. *Biol Psychiatry Cogn Neurosci Neuroimaging*.
848 2017;2(2):158-169. doi:10.1016/j.bpsc.2016.10.002

849 34. Nakayama D, Baraki Z, Onoue K, Ikegaya Y, Matsuki N, Nomura H. Frontal association
850 cortex is engaged in stimulus integration during associative learning. *Curr Biol*.
851 2015;25(1):117-123. doi:10.1016/j.cub.2014.10.078

852 35. Bekinschtein P, Cammarota M, Igaz LM, Bevilaqua LRM, Izquierdo I, Medina JH.
853 Persistence of Long-Term Memory Storage Requires a Late Protein Synthesis- and BDNF-
854 Dependent Phase in the Hippocampus. *Neuron*. 2007;53(2):261-277.
855 doi:10.1016/j.neuron.2006.11.025

856 36. Blundell J, Kouser M, Powell CM. Systemic inhibition of mammalian target of rapamycin
857 inhibits fear memory reconsolidation. *Neurobiol Learn Mem*. 2008;90(1):28-35.
858 doi:10.1016/j.nlm.2007.12.004

859 37. Mac Callum PE, Hebert M, Adamec RE, Blundell J. Systemic inhibition of mTOR kinase via
860 rapamycin disrupts consolidation and reconsolidation of auditory fear memory.
861 *Neurobiol Learn Mem*. 2014;112:176-185. doi:10.1016/j.nlm.2013.08.014

862 38. Gould TJ, Wilkinson DS, Yildirim E, Poole RLF, Leach PT, Simmons SJ. Nicotine shifts the
863 temporal activation of hippocampal protein kinase A and extracellular signal-regulated
864 kinase 1/2 to enhance long-term, but not short-term, hippocampus-dependent memory.
865 *Neurobiol Learn Mem*. 2014;109:151-159. doi:10.1016/j.nlm.2014.01.009

866 39. Kwapis JL, Jarome TJ, Lee JL, Helmstetter FJ. The retrosplenial cortex is involved in the
867 formation of memory for context and trace fear conditioning. *Neurobiol Learn Mem*.
868 2015;123:110-116. doi:10.1016/j.nlm.2015.06.007

869 40. Jarome TJ, Ferrara NC, Kwapis JL, Helmstetter FJ. CaMKII regulates proteasome
870 phosphorylation and activity and promotes memory destabilization following retrieval.
871 *Neurobiol Learn Mem*. 2016;128:103-109. doi:10.1016/j.nlm.2016.01.001

872 41. Signor C, Temp FR, Mello CF, et al. Intrahippocampal infusion of spermidine improves
873 memory persistence: Involvement of protein kinase A. *Neurobiol Learn Mem*.
874 2016;131:18-25. doi:10.1016/j.nlm.2016.03.003

875 42. Bhattacharya S, Kimble W, Buabeid M, et al. Altered AMPA receptor expression plays an
876 important role in inducing bidirectional synaptic plasticity during contextual fear memory
877 reconsolidation. *Neurobiol Learn Mem*. 2017;139:98-108. doi:10.1016/j.nlm.2016.12.013

878 43. Schmidt SD, Furini CRG, Zinn CG, et al. Modulation of the consolidation and
879 reconsolidation of fear memory by three different serotonin receptors in hippocampus.
880 *Neurobiol Learn Mem*. 2017;142:48-54. doi:10.1016/j.nlm.2016.12.017

881 44. Kumar T, Jha SK. Influence of cued-fear conditioning and its impairment on NREM sleep.
882 *Neurobiol Learn Mem*. 2017;144:155-165. doi:10.1016/j.nlm.2017.07.008

883 45. Song Z, Chen H, Xu W, Wu S, Zhu G. Basolateral amygdala calpain is required for
884 extinction of contextual fear-memory. *Neurobiol Learn Mem*. 2018;155:180-188.
885 doi:10.1016/j.nlm.2018.08.004

886 46. Maciejak P, Szyndler J, Lehner M, et al. The differential effects of protein synthesis
887 inhibition on the expression and reconsolidation of pentylenetetrazole kindled seizures.
888 *Epilepsy Behav*. 2010;18(3):193-200. doi:10.1016/j.yebeh.2010.04.005

889 47. Abel T, Nguyen PV, Barad M, Deuel TAS, Kandel ER, Bourtchouladze R. Genetic
890 demonstration of a role for PKA in the late phase of LTP and in hippocampus-based long-
891 term memory. *Cell*. 1997;88(5):615-626. doi:10.1016/S0092-8674(00)81904-2

892 48. Barrientos RM, Higgins EA, Sprunger DB, Watkins LR, Rudy JW, Maier SF. Memory for
893 context is impaired by injecting anisomycin into dorsal hippocampus following context
894 exploration. *Behav Brain Res*. 2002;134(1-2):299-306. doi:10.1016/S0166-
895 4328(02)00045-1

896 49. Debiec J, LeDoux JE, Nader K. Cellular and systems reconsolidation in the hippocampus.
897 *Neuron*. 2002;36(3):527-538. doi:10.1016/S0896-6273(02)01001-2

898 50. Huff NC, Rudy JW. The amygdala modulates hippocampus-dependent context memory
899 formation and stores cue-shock associations. *Behav Neurosci*. 2004;118(1):53-62.
900 doi:10.1037/0735-7044.118.1.53

901 51. Biedenkapp JC, Rudy JW. Context memories and reactivation: constraints on the
902 reconsolidation hypothesis. *Behav Neurosci*. 2004;118(5):956-964. doi:10.1037/0735-
903 7044.118.5.956

904 52. Rudy JW, Matus-Amat P. The ventral hippocampus supports a memory representation of
905 context and contextual fear conditioning: implications for a unitary function of the
906 hippocampus. *Behav Neurosci*. 2005;119(1):154-163. doi:10.1037/0735-7044.119.1.154

907 53. Kochli DE, Campbell TL, Hollingsworth EW, et al. Combined administration of MK-801 and
908 cycloheximide produces a delayed potentiation of fear discrimination memory extinction.
909 *Behav Neurosci*. 2018;132(2):99-105. doi:10.1037/bne0000232

910 54. Trent S, Barnes P, Hall J, Thomas KL. Rescue of long-term memory after reconsolidation
911 blockade. *Nat Commun*. 2015;6:1-7. doi:10.1038/ncomms8897

912 55. Dębiec J, Díaz-Mataix L, Bush DEA, Doyère V, Ledoux JE. The amygdala encodes specific
913 sensory features of an aversive reinforcer. *Nat Neurosci*. 2010;13(5):536-537.
914 doi:10.1038/nn.2520

915 56. Mamou CB, Gamache K, Nader K. NMDA receptors are critical for unleashing
916 consolidated auditory fear memories. *Nat Neurosci*. 2006;9(10):1237-1239.
917 doi:10.1038/nn1778

918 57. Nomura H, Matsuki N. Ethanol enhances reactivated fear memories.
919 *NEUROPSYCHOPHARMACOLOGY*. 2008;33(12):2912-2921. doi:10.1038/npp.2008.13

920 58. Yang CH, Huang CC, Hsu KS. Generalization of fear inhibition by disrupting hippocampal
921 protein synthesis-dependent reconsolidation process. *Neuropsychopharmacol Off Publ*
922 *Am Coll Neuropsychopharmacol*. 2011;36(10):1992-2008. doi:10.1038/npp.2011.87

923 59. Haubrich J, Crestani AP, Cassini LF, et al. Reconsolidation Allows Fear Memory to Be
924 Updated to a Less Aversive Level through the Incorporation of Appetitive Information.
925 *NEUROPSYCHOPHARMACOLOGY*. 2015;40(2):315-326. doi:10.1038/npp.2014.174

926 60. Chai N, Liu JF, Xue YX, et al. Delayed Noradrenergic Activation in the Dorsal Hippocampus
927 Promotes the Long-Term Persistence of Extinguished Fear.
928 *NEUROPSYCHOPHARMACOLOGY*. 2014;39(8):1933-1945. doi:10.1038/npp.2014.42

929 61. Awad W, Ferreira G, Maroun M. Dissociation of the Role of Infralimbic Cortex in Learning
930 and Consolidation of Extinction of Recent and Remote Aversion Memory.
931 *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol*. 2015;40(11):2566-
932 2575. doi:10.1038/npp.2015.103

933 62. Jarome TJ, Ferrara NC, Kwapis JL, Helmstetter FJ. Contextual Information Drives the
934 Reconsolidation-Dependent Updating of Retrieved Fear Memories.
935 *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol*. 2015;40(13):3044-
936 3052. doi:10.1038/npp.2015.161

937 63. Kwapis JL, Jarome TJ, Ferrara NC, Helmstetter FJ. Updating Procedures can Reorganize
938 the Neural Circuit Supporting a Fear Memory. *Neuropsychopharmacol Off Publ Am Coll*
939 *Neuropsychopharmacol*. Published online January 2017. doi:10.1038/npp.2017.23

940 64. Haubrich J, Machado A, Boos FZ, et al. Enhancement of extinction memory by
941 pharmacological and behavioral interventions targeted to its reactivation. *Sci Rep*.
942 2017;7(1):1-11. doi:10.1038/s41598-017-11261-6

943 65. Lattal KM, Abel T. Behavioral impairments caused by injections of the protein synthesis
944 inhibitor anisomycin after contextual retrieval reverse with time. *Proc Natl Acad Sci U S*
945 *A*. 2004;101(13):4667-4672. doi:10.1073/pnas.0306546101

946 66. Dębiec J, Doyère V, Nader K, LeDoux JE. Directly reactivated, but not indirectly
947 reactivated, memories undergo reconsolidation in the amygdala. *Proc Natl Acad Sci U S*
948 *A*. 2006;103(9):3428-3433. doi:10.1073/pnas.0507168103

949 67. Ballarini F, Moncada D, Martinez MC, Alen N, Viola H. Behavioral tagging is a general
950 mechanism of long-term memory formation. *Proc Natl Acad Sci U S A*.
951 2009;106(34):14599-14604. doi:10.1073/pnas.0907078106

952 68. Poulos AM, Ponnusamy R, Dong HW, Fanselow MS. Compensation in the neural circuitry
953 of fear conditioning awakens learning circuits in the bed nuclei of the stria terminalis.
954 *Proc Natl Acad Sci U S A*. 2010;107(33):14881-14886. doi:10.1073/pnas.1005754107

955 69. Hoeffer CA, Cowansage KK, Arnold EC, et al. Inhibition of the interactions between
956 eukaryotic initiation factors 4E and 4G impairs long-term associative memory
957 consolidation but not reconsolidation. *Proc Natl Acad Sci U S A*. 2011;108(8):3383-3388.
958 doi:10.1073/pnas.1013063108

959 70. Radwanska K, Medvedev NI, Pereira GS, et al. Mechanism for long-term memory
960 formation when synaptic strengthening is impaired. *Proc Natl Acad Sci U S A*.
961 2011;108(45):18471-18475. doi:10.1073/pnas.1109680108

962 71. de Carvalho Myskiw J, Furini CRG, Schmidt B, Ferreira F, Izquierdo I. Extinction learning,
963 which consists of the inhibition of retrieval, can be learned without retrieval. *Proc Natl*
964 *Acad Sci U S A*. 2015;112(2):E230-3. doi:10.1073/pnas.1423465112

965 72. Roy DS, Muralidhar S, Smith LM, Tonegawa S. Silent memory engrams as the basis for
966 retrograde amnesia. *Proc Natl Acad Sci U S A*. 2017;114(46):E9972-E9979.
967 doi:10.1073/pnas.1714248114

968 73. Kwapis JL, Jarome TJ, Schiff JC, Helmstetter FJ. Memory consolidation in both trace and
969 delay fear conditioning is disrupted by intra-amygdala infusion of the protein synthesis
970 inhibitor anisomycin. *Learn Mem Cold Spring Harb N*. 2011;18(11):728-732.
971 doi:10.1101/lm.023945.111

972 74. Jarome TJ, Kwapis JL, Werner CT, Parsons RG, Gafford GM, Helmstetter FJ. The timing of
973 multiple retrieval events can alter GluR1 phosphorylation and the requirement for
974 protein synthesis in fear memory reconsolidation. *Learn Mem*. 2012;19(7):300-306.
975 doi:10.1101/lm.024901.111

976 75. Einarsson EÖ, Nader K. Involvement of the anterior cingulate cortex in formation,
977 consolidation, and reconsolidation of recent and remote contextual fear memory. *Learn*
978 *Mem*. 2012;19(10):449-452. doi:10.1101/lm.027227.112

979 76. Nakayama D, Yamasaki Y, Matsuki N, Nomura H. Post-retrieval late process contributes
980 to persistence of reactivated fear memory. *Learn Mem*. 2013;20(6):307-310.
981 doi:10.1101/lm.029660.112

982 77. Remaud J, Ceccom J, Carponcy J, et al. Anisomycin injection in area CA3 of the
983 hippocampus impairs both short-term and long-term memories of contextual fear. *Learn*
984 *Mem*. 2014;21(6):311-315. doi:10.1101/lm.033969.113

985 78. Kochli DE, Thompson EC, Fricke EA, Postle AF, Quinn JJ. The amygdala is critical for trace,
986 delay, and contextual fear conditioning. *Learn Mem*. 2015;22(2):92-100.
987 doi:10.1101/lm.034918.114

988 79. Almeida-Corrêa S, Moulin TC, Carneiro CFD, Gonçalves MMC, Junqueira LS, Amaral OB.
989 Calcineurin inhibition blocks within-, but not between-session fear extinction in mice.
990 *Learn Mem*. 2015;22(3):159-169. doi:10.1101/lm.037770.114

991 80. Girardi BA, Ribeiro DA, Signor C, et al. Spermidine-induced improvement of
992 reconsolidation of memory involves calcium-dependent protein kinase in rats. *Learn*
993 *Mem*. 2016;23(1):21-28. doi:10.1101/lm.039396.115

994 81. Ferrara NC, Cullen PK, Pullins SP, Rotondo EK, Helmstetter FJ. Input from the medial
995 geniculate nucleus modulates amygdala encoding of fear memory discrimination. *Learn*
996 *Mem*. 2017;24(9):414-421. doi:10.1101/lm.044131.116

997 82. Viosca J, De Armentia ML, Jancic D, Barco A. Enhanced CREB-dependent gene expression
998 increases the excitability of neurons in the basal amygdala and primes the consolidation
999 of contextual and cued fear memory. *Learn Mem*. 2009;16(3):193-197.
1000 doi:10.1101/lm.1254209

1001 83. Stafford JM, Lattal KM. Direct comparisons of the size and persistence of anisomycin-
1002 induced consolidation and reconsolidation deficits. *Learn Mem*. 2009;16(8):494-503.
1003 doi:10.1101/lm.1452209

1004 84. Ageta H, Ikegami S, Miura M, et al. Activin plays a key role in the maintenance of long-
1005 term memory and late-LTP. *Learn Mem*. 2010;17(4):176-185. doi:10.1101/lm.16659010

1006 85. Frankland PW, Ding HK, Takahashi E, Suzuki A, Kida S, Silva AJ. Stability of recent and
1007 remote contextual fear memory. *Learn Mem*. 2006;13(4):451-457.
1008 doi:10.1101/lm.183406

1009 86. Suzuki A, Mukawa T, Tsukagoshi A, Frankland PW, Kida S. Activation of LVGCCs and CB1
1010 receptors required for destabilization of reactivated contextual fear memories. *Learn*
1011 *Mem*. 2008;15(6):426-433. doi:10.1101/lm.888808

1012 87. Bal NV, Rysakova MP, Vinarskaya AK, Ivanova V, Zuzina AB, Balaban PM. Cued memory
1013 reconsolidation in rats requires nitric oxide. *Eur J Neurosci*. 2017;45(5):643-647.
1014 doi:10.1111/ejn.13503

1015 88. Duvarci S, Nader K, LeDoux JE. Activation of extracellular signal-regulated kinase-
1016 mitogen-activated protein kinase cascade in the amygdala is required for memory
1017 reconsolidation of auditory fear conditioning. *Eur J Neurosci*. 2005;21(1):283-289.
1018 doi:10.1111/j.1460-9568.2004.03824.x

1019 89. Parsons RG, Gafford GM, Baruch DE, Riedner BA, Helmstetter FJ. Long-term stability of
1020 fear memory depends on the synthesis of protein but not mRNA in the amygdala. *Eur J*
1021 *Neurosci*. 2006;23(7):1853-1859. doi:10.1111/j.1460-9568.2006.04723.x

1022 90. Duvarci S, Mamou C ben, Nader K. Extinction is not a sufficient condition to prevent fear
1023 memories from undergoing reconsolidation in the basolateral amygdala. *Eur J Neurosci*.
1024 2006;24(1):249-260. doi:10.1111/j.1460-9568.2006.04907.x

1025 91. Sacchetti B, Sacco T, Strata P. Reversible inactivation of amygdala and cerebellum but not
1026 perirhinal cortex impairs reactivated fear memories. *Eur J Neurosci*. 2007;25(9):2875-
1027 2884. doi:10.1111/j.1460-9568.2007.05508.x

1028 92. Lee SH, Choi JH, Lee N, et al. Synaptic protein degradation underlies destabilization of
1029 retrieved fear memory. *Science*. 2008;319(5867):1253-1256.
1030 doi:10.1126/science.1150541

1031 93. Blum S, Runyan JD, Dash PK. Inhibition of prefrontal protein synthesis following recall
1032 does not disrupt memory for trace fear conditioning. *BMC Neurosci*. 2006;7:67.
1033 doi:10.1186/1471-2202-7-67

1034 94. Kishioka A, Uemura T, Fukushima F, Mishina M. Consolidation of auditory fear memories
1035 formed by weak unconditioned stimuli requires NMDA receptor activation and de novo
1036 protein synthesis in the striatum. *Mol Brain*. 2013;6:17. doi:10.1186/1756-6606-6-17

1037 95. Barnes P, Thomas KL. Proteolysis of proBDNF Is a Key Regulator in the Formation of
1038 Memory. Bartolomucci A, ed. *PLoS ONE*. 2008;3(9):e3248.
1039 doi:10.1371/journal.pone.0003248

1040 96. Apergis-Schoute AM, Dębiec J, Doyère V, LeDoux JE, Schafe GE. Auditory fear
1041 conditioning and long-term potentiation in the lateral amygdala require ERK/MAP kinase
1042 signaling in the auditory thalamus: A role for presynaptic plasticity in the fear system. *J*
1043 *Neurosci*. 2005;25(24):5730-5739. doi:10.1523/JNEUROSCI.0096-05.2005

1044 97. Lopez J, Gamache K, Schneider R, Nader K. Memory retrieval requires ongoing protein
1045 synthesis and NMDA receptor activity-mediated AMPA receptor trafficking. *J Neurosci*.
1046 2015;35(6):2465-2475. doi:10.1523/JNEUROSCI.0735-14.2015

1047 98. Santini E, Ge H, Ren K, Peña De Ortiz S, Quirk GJ. Consolidation of fear extinction requires
1048 protein synthesis in the medial prefrontal cortex. *J Neurosci*. 2004;24(25):5704-5710.
1049 doi:10.1523/JNEUROSCI.0786-04.2004

1050 99. Kim J, Song B, Hong I, et al. Reactivation of Fear Memory Renders Consolidated Amygdala
1051 Synapses Labile. *J Neurosci*. 2010;30(28):9631-9640. doi:10.1523/JNEUROSCI.0940-
1052 10.2010

1053 100. Holehonnur R, Phensy AJ, Kim LJ, et al. Increasing the GluN2A/GluN2B ratio in neurons of
1054 the mouse basal and lateral amygdala inhibits the modification of an existing fear
1055 memory trace. *J Neurosci*. 2016;36(36):9490-9504. doi:10.1523/JNEUROSCI.1743-
1056 16.2016

1057 101. Cai WH, Blundell J, Han J, Greene RW, Powell CM. Postreactivation glucocorticoids impair
1058 recall of established fear memory. *J Neurosci*. 2006;26(37):9560-9566.
1059 doi:10.1523/JNEUROSCI.2397-06.2006

1060 102. Duvarci S, Nader K. Characterization of fear memory reconsolidation. In: *Journal of*
1061 *Neuroscience*. Vol 24. Society for Neuroscience; 2004:9269-9275.
1062 doi:10.1523/JNEUROSCI.2971-04.2004

1063 103. Wilensky AE, Schafe GE, Kristensen MP, LeDoux JE. Rethinking the fear circuit: The
1064 central nucleus of the amygdala is required for the acquisition, consolidation, and
1065 expression of pavlovian fear conditioning. *J Neurosci*. 2006;26(48):12387-12396.
1066 doi:10.1523/JNEUROSCI.4316-06.2006

1067 104. Mamiya N, Fukushima H, Suzuki A, et al. Brain region-specific gene expression activation
1068 required for reconsolidation and extinction of contextual fear memory. *J Neurosci*.
1069 2009;29(2):402-413. doi:10.1523/JNEUROSCI.4639-08.2009

1070 105. von Hertzen LSJ. Memory Reconsolidation Engages Only a Subset of Immediate-Early
1071 Genes Induced during Consolidation. *J Neurosci*. 2005;25(8):1935-1942.
1072 doi:10.1523/JNEUROSCI.4707-04.2005

1073 106. Fischer A, Sananbenesi F, Schrick C, Spiess J, Radulovic J. Distinct Roles of Hippocampal
1074 De Novo Protein Synthesis and Actin Rearrangement in Extinction of Contextual Fear. *J*
1075 *Neurosci*. 2004;24(8):1962-1966. doi:10.1523/JNEUROSCI.5112-03.2004

1076 107. Reis DS, Jarome TJ, Helmstetter FJ. Memory formation for trace fear conditioning
1077 requires ubiquitin-proteasome mediated protein degradation in the prefrontal cortex.
1078 *Front Behav Neurosci*. 2013;7(OCT):150. doi:10.3389/fnbeh.2013.00150

1079 108. Kwak C, Choi JH, Bakes JT, Lee K, Kaang BK. Effect of intensity of unconditional stimulus
1080 on reconsolidation of contextual fear memory. *Korean J Physiol Pharmacol Off J Korean*
1081 *Physiol Soc Korean Soc Pharmacol*. 2012;16(5):293-296. doi:10.4196/kjpp.2012.16.5.293

1082 109. Stiedl O, Palve M, Radulovic J, Birkenfeld K, Spiess J. Differential impairment of auditory
1083 and contextual fear conditioning by protein synthesis inhibition in C57BL/6N mice. *Behav*
1084 *Neurosci*. 1999;113(3):496-506. doi:10.1037/0735-7044.113.3.496

1085 110. Bourtchouladze R, Abel T, Berman N, Gordon R, Lapidus K, Kandel ER. Different training
1086 procedures recruit either one or two critical periods for contextual memory
1087 consolidation, each of which requires protein synthesis and PKA. *Learn Mem*. 1998;5(4-
1088 5):365-374. doi:10.1101/lm.5.4.365

1089 111. Schafe GE, LeDoux JE. Memory consolidation of auditory pavlovian fear conditioning
1090 requires protein synthesis and protein kinase A in the amygdala. *J Neurosci Off J Soc*
1091 *Neurosci*. 2000;20(18):RC96-RC96. doi:10.1523/jneurosci.20-18-j0003.2000

1092 112. Lattal KM, Abel T. Different requirements for protein synthesis in acquisition and
1093 extinction of spatial preferences and context-evoked fear. *J Neurosci*. 2001;21(15):5773-
1094 5780. doi:10.1523/jneurosci.21-15-05773.2001

1095 113. Scharf MT, Woo NH, Lattal KM, Young JZ, Nguyen PV, Abel T. Protein Synthesis Is
1096 Required for the Enhancement of Long-Term Potentiation and Long-Term Memory by
1097 Spaced Training. *J Neurophysiol*. 2002;87(6):2770-2777. doi:10.1152/jn.2002.87.6.2770

1098 114. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and
1099 meta-analyses: The PRISMA statement. *PLoS Med*. 2009;6(7).
1100 doi:10.1371/journal.pmed.1000097

1101 115. Davis HP, Squire LR. Protein synthesis and memory: A review. *Psychol Bull*.
1102 1984;96(3):518-559. doi:10.1037/0033-2909.96.3.518

1103 116. Nader K. Memory traces unbound. *Trends Neurosci*. 2003;26(2):65-72.
1104 doi:10.1016/S0166-2236(02)00042-5

1105 117. Vianna MR, Szapiro G, McGaugh JL, Medina JH, Izquierdo I. Retrieval of memory for fear-
1106 motivated training initiates extinction requiring protein synthesis in the rat hippocampus.
1107 *Proc Natl Acad Sci U S A*. 2001;98(21):12251-12254. doi:10.1073/pnas.211433298

1108 118. Berman DE, Dudai Y. Memory extinction, learning anew, and learning the new:
1109 Dissociations in the molecular machinery of learning in cortex. *Science*.
1110 2001;291(5512):2417-2419. doi:10.1126/science.1058165

1111 119. Tiunova AA, Anokhin KV, Rose SPR. Two critical periods of protein and glycoprotein
1112 synthesis in memory consolidation for visual categorization learning in chicks. *Learn*
1113 *Mem*. 1998;4(5):401-410. doi:10.1101/lm.4.5.401

1114 120. Pedreira ME, Pérez-Cuesta LM, Maldonado H. Mismatch between what is expected and
1115 what actually occurs triggers memory reconsolidation or extinction. *Learn Mem*.
1116 2004;11(5):579-585. doi:10.1101/lm.76904

1117 121. Sevenster D, Beckers T, Kindt M. Prediction error demarcates the transition from
1118 retrieval, to reconsolidation, to new learning. *Learn Mem*. 2014;21(11):580-584.
1119 doi:10.1101/lm.035493.114

1120 122. Alfei JM, Monti RIF, Molina VA, Bueno AM, Urcelay GP. Prediction error and trace
1121 dominance determine the fate of fear memories after post-training manipulations. *Learn*
1122 *Mem*. 2015;22(8):385-400. doi:10.1101/lm.038513.115

1123 123. Eisenberg M, Kobilo T, Berman DE, Dudai Y. Stability of retrieved memory: inverse
1124 correlation with trace dominance. *Science*. 2003;301(5636):1102-1104.
1125 doi:10.1126/science.1086881

1126 124. Wang SH, De Oliveira Alvares L, Nader K. Cellular and systems mechanisms of memory
1127 strength as a constraint on auditory fear reconsolidation. *Nat Neurosci*. 2009;12(7):905-
1128 912. doi:10.1038/nn.2350

1129 125. Abrari K, Rashidy-Pour A, Semnanian S, Fathollahi Y. Administration of corticosterone
1130 after memory reactivation disrupts subsequent retrieval of a contextual conditioned fear
1131 memory: Dependence upon training intensity. *Neurobiol Learn Mem*. 2008;89(2):178-
1132 184. doi:10.1016/j.nlm.2007.07.005

1133 126. Milekic MH, Alberini CM. Temporally graded requirement for protein synthesis following
1134 memory reactivation. *Neuron*. 2002;36(3):521-525. doi:10.1016/S0896-6273(02)00976-5

1135 127. Eisenberg M, Dudai Y. Reconsolidation of fresh, remote, and extinguished fear memory in
1136 medaka: old fears don't die. *Eur J Neurosci*. 2004;20(12):3397-3403. doi:10.1111/j.1460-
1137 9568.2004.03818.x

1138 128. Rotondo F, Biddle K, Chen J, Ferencik J, d'Esneval M, Milton AL. Lack of Effect of
1139 Propranolol on the Reconsolidation of Conditioned Fear Memory due to a Failure to
1140 Engage Memory Destabilisation. *Neuroscience*. 2022;480:9-18.
1141 doi:10.1016/j.neuroscience.2021.11.008

1142 129. Soeter M, Kindt M. High Trait Anxiety: A Challenge for Disrupting Fear Memory
1143 Reconsolidation. *PLOS ONE*. 2013;8(11). doi:10.1371/journal.pone.0075239

1144 130. Schimanski LA, Nguyen PV. Multidisciplinary approaches for investigating the
1145 mechanisms of hippocampus-dependent memory: A focus on inbred mouse strains.
1146 *Neurosci Biobehav Rev*. 2004;28(5):463-483. doi:10.1016/j.neubiorev.2004.04.002

1147 131. Macleod MR, Lawson McLean A, Kyriakopoulou A, et al. Risk of Bias in Reports of In Vivo
1148 Research: A Focus for Improvement. *PLOS Biol*. 2015;13(10):e1002273.
1149 doi:10.1371/journal.pbio.1002273

1150 132. Carneiro CFD, Moulin TC, Macleod MR, Amaral OB. Effect size and statistical power in the
1151 rodent fear conditioning literature - a systematic review. *PLoS ONE*. 2018.

1152 133. Tannenbaum C, Ellis RP, Eyssel F, Zou J, Schiebinger L. Sex and gender analysis improves
1153 science and engineering. *Nature*. 2019;575(7781):137-146. doi:10.1038/s41586-019-
1154 1657-6

1155 134. Clayton JA, Collins FS. Policy: NIH to balance sex in cell and animal studies. *Nature*.
1156 2014;509(7500):282-283.

1157 135. Amaral OB, Neves K. Reproducibility: expect less of the scientific paper. *Nature*.
1158 2021;597(7876):329-331. doi:10.1038/d41586-021-02486-7

1159 136. Bannach-Brown A, Hair K, Bahor Z, Soliman N, MacLeod M, Liao J. Technological
1160 advances in preclinical meta-research. *BMJ Open Sci*. 2021;5(1):e100131.
1161 doi:10.1136/bmjos-2020-100131

1162