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 multiple 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 (systemic, hippocampus, amygdala, and cerebral ventricles) and target sessions (training, reconsolidation and extinction) to evaluate the effect of the intervention in various scenarios. Heterogeneity was further investigated by uni- and multivariable meta-regression using 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 injection timing for training interventions and reexposure duration for reactivation interventions. However, other factors proposed as boundary conditions for reconsolidation, such as training strength and memory age, were not associated with effect size across the literature. Additionally, our findings suggest some underexplored associations, such as a larger susceptibility of cued fear conditioning to reconsolidation when compared to the contextual version of the task. 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.

Keywords

Preclinical meta-analysis; heterogeneity; fear conditioning; protein synthesis

Corresponding author:

Olavo B. Amaral, olavo@bioqmed.ufrj.br

Instituto de Bioquímica Médica Leopoldo de Meis
Universidade Federal do Rio de Janeiro
Av. Carlos Chagas Filho 373, Bloco E-38.
Rio de Janeiro, RJ, Brazil. 22291-050.

Introduction

Over the last decades, biomedical research has grown, accumulated data, developed new technologies and published growing numbers of articles; however, low reproducibility rates in some areas have shed doubt over its reliability (Amaral et al., 2019; Errington et al., 2021; Open Science Collaboration, 2015). Industry figures for the replicability of basic research have been as low as 11% and 21% (Begley & Ellis, 2012; Prinz et al., 2011). In experimental neurology, a review of replications in spinal cord injury research reported that only 30% of them were at least partly successful (Steward et al., 2012). In basic cancer biology, a 6-year systematic effort reported that 46% of experiments from highly cited articles were reproduced by most of the study's criteria, with effect sizes being 85% smaller in replications than in the original studies (Errington et al., 2021).

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 and explore sources of heterogeneity among studies (Vesterinen et al., 2014). 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 (França & Monserrat, 2018; Stanley & Spence, 2014).

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 many control experiments testing boundary conditions (i.e. changing protocol parameters in order to establish the necessary conditions for an effect to be observed). This makes it much harder to fit these studies into traditional frameworks for describing research questions, such as Population-Intervention-Comparator-Outcome (PICO). It also leads to multiple experiments published in individual articles, as well as many articles by the same research groups, violating the assumption of independence between different results when synthesizing data.

While concerns over reproducibility have received greater attention in some 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 a negative unconditioned stimulus (US; for example, footshocks) (Wotjak, 2019). 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, as well as to study reconsolidation and extinction phenomena, where the fearful association is changed or replaced by a non-fearful one.

After a series of failed attempts to replicate findings on reconsolidation of fear conditioning in rodents (Luyten et al., 2021; Schroyens et al., 2019), Schroyens *et al*. found evidence of publication bias in the field by reviewing the literature and contacting relevant authors (Schroyens et al., 2021). In their sample of studies, 80% of published findings yielded statistically significant results, in contrast to only 20% of unpublished

findings they had access to (Schroyens et al., 2021). The failed replications also raise the possibility that differences in protocols might determine the magnitude of the outcome. This suggests a necessity to determine the robustness of the effect of amnestic agents in different conditions, as well as to systematically investigate moderators that may explain heterogeneity in results.

In this work, we use a classic intervention – 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 doing this, we aim to investigate whether meta-analytic methods can detect well-established findings – such as the effect of protein synthesis inhibition on consolidation – and whether it can shed light on more controversial topics – 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 behavioural 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.

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. Multiple 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 (version 4.0.4, (R Core Team, 2021). 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 (Moulin & Amaral, 2020). Briefly, the list of authors was obtained for all articles included in our dataset to create a co-authorship graph using the Louvain method (Blondel et al., 2008). 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). 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 (Egger et al., 1997) and trim-and-fill analyses (Duval & Tweedie, 2000a, 2000b). As recommended by the authors of the method (Duval, 2005), 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 (Ioannidis & Trikalinos, 2007) 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 (Abate et al., 2018; Abel et al., 1997; Ageta et al., 2010; Almeida-Corrêa et al., 2015; Apergis-Schoute et al., 2005; Awad et al., 2015; Bal et al., 2017; Ballarini et al., 2009; Barnes & Thomas, 2008; Barrientos et al., 2002; Bekinschtein et al., 2007; Bhattacharya et al., 2017; Biedenkapp & Rudy, 2004; Blum et al., 2006; Blundell et al., 2008; Bourtchouladze et al., 1998; Cai et al., 2006; Chai et al., 2014; de Carvalho Myskiw et al., 2015; Debiec et al., 2002, 2006, 2010; Duvarci et al., 2005, 2006; Duvarci & Nader, 2004; Einarsson & Nader, 2012; Ferrara et al., 2017; Fischer et al., 2004; Frankland et al., 2004, 2006; Girardi et al., 2016; Gould et al., 2014; Haubrich et al., 2015, 2017; Hoeffer et al., 2011; Holehonnur et al., 2016; Huff & Rudy, 2004; Jarome et al., 2012, 2015, 2016; Kim et al., 2010; Kishioka et al., 2013; Kochli et al., 2015, 2018; Kumar & Jha, 2017; Kwak et al., 2012; Kwapis et al., 2011, 2015, 2017; Lattal & Abel, 2001, 2004; Lee et al., 2008; Lopez et al., 2015; Mac Callum et al., 2014; Maciejak et al., 2010; Mamiya et al., 2009; Mamou et al., 2006; Motanis & Maroun, 2012; Nader et al., 2000; Nakayama et al., 2013, 2015; Nomura & Matsuki, 2008; Parsons et al., 2006; Casanova et al., 2016; Poulos et al., 2010; Radwanska et al., 2011; Reis et al., 2013; Remaud et al., 2014; Rizzo et al., 2017; Roy et al., 2017; Rudy & Matus-Amat, 2005; Runyan & Dash, 2005; Sacchetti et al., 2007; Santini et al., 2004; Schafe & LeDoux, 2000; Scharf et al., 2002; Schmidt et al., 2017; Signor et al., 2016; Song et al., 2018; Stafford & Lattal, 2009; Stiedl et al., 1999; Suzuki et al., 2004, 2008;

Trent et al., 2015; Viosca et al., 2009; von Hertzen, 2005; Wanisch et al., 2005; Wilensky et al., 2006; Jin et al., 2007; Yang et al., 2011; Yim et al., 2006). **Figure 1** presents a PRISMA flowchart for the inclusion and exclusion of articles (Moher et al., 2009).

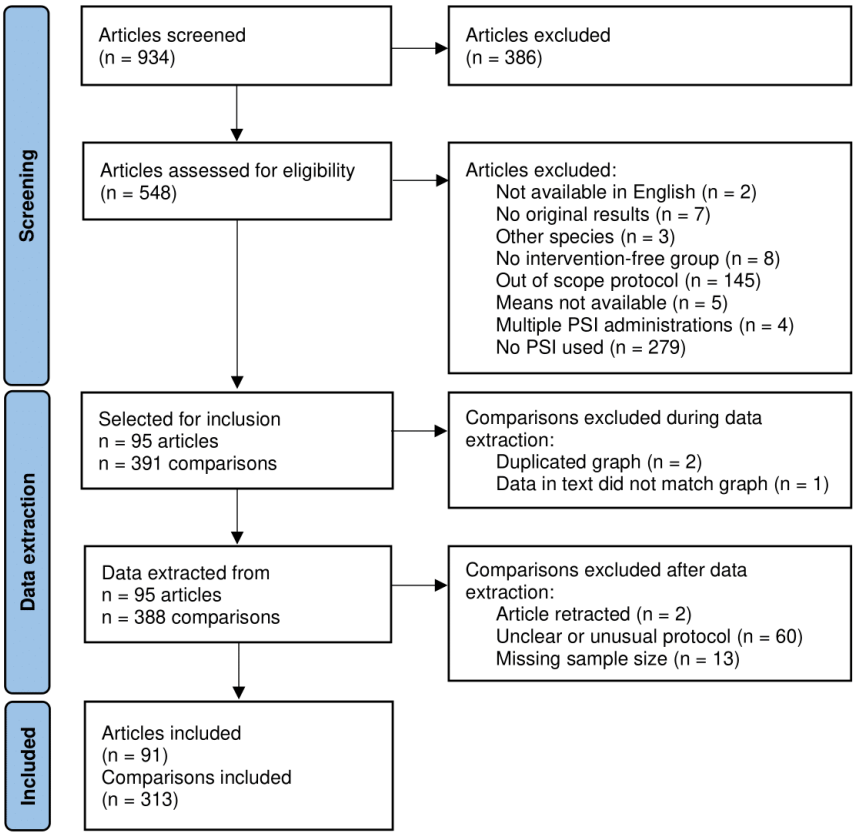

Figure 1 - Flowchart of screening stages. In the first stage of screening (n=934), two reviewers assessed titles and abstracts. If one reviewer included the article, it was selected for full-text screening. During this stage we did not record reasons for exclusion. The second screening stage (n=548) 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 general features (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. To assess risk of bias, we recorded the reporting of random allocation of animals to groups, blinded or automated outcome assessment, sample size calculation, statements of compliance with regulatory requirements and conflicts of interest (**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**, with the full data available 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]	$6.4x10^{-25}$	82.8%	$1.5x10^{-38}$
Training, amygdala	38	-20.2 [-26.1, -14.3]	$1.5x10^{-11}$	66.6%	$1.5x10^{-10}$
Training, hippocampus	41	-14.1 [-18.9, -9.3]	$9.0x10^{-9}$	87.3%	$1.9x10^{-23}$
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]	$4.2x10^{-16}$	83.4%	$4.4x10^{-35}$
Reconsolidation, amygdala	47	-20.1 [-25.2, -15.0]	$1.4x10^{-14}$	78.8%	$5.5x10^{-26}$
Reconsolidation, hippocampus	20	-19.3 [-27.8, -10.9]	$7.8x10^{-6}$	89.6%	$1.9x10^{-34}$
Reconsolidation, i.c.v.	4	-12.5 [-43.9, 18.9]	0.435	88.5%	$9.0x10^{-4}$
Extinction, systemic	4	28.0 [21.4, 34.6]	$1.2x10^{-16}$	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]	$3.3x10^{-6}$	91.7%	$1.3x10^{-66}$
Reactivation, amygdala	56	-18.1 [-23.2, -13.0]	$2.9x10^{-12}$	82.4%	$2.3x10^{-43}$
Reactivation, hippocampus	30	-13.2 [-21.3, -5.2]	0.001	90.7%	$1.9x10^{-59}$
Reactivation, i.c.v.	6	0.9 [-21.7, 23.5]	0.939	91.6%	$2.8x10^{-5}$
Complete	313	-15.4 [-17.4, -13.3]	$6.0x10^{-48}$	87.0%	$6.5x10^{-284}$

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.

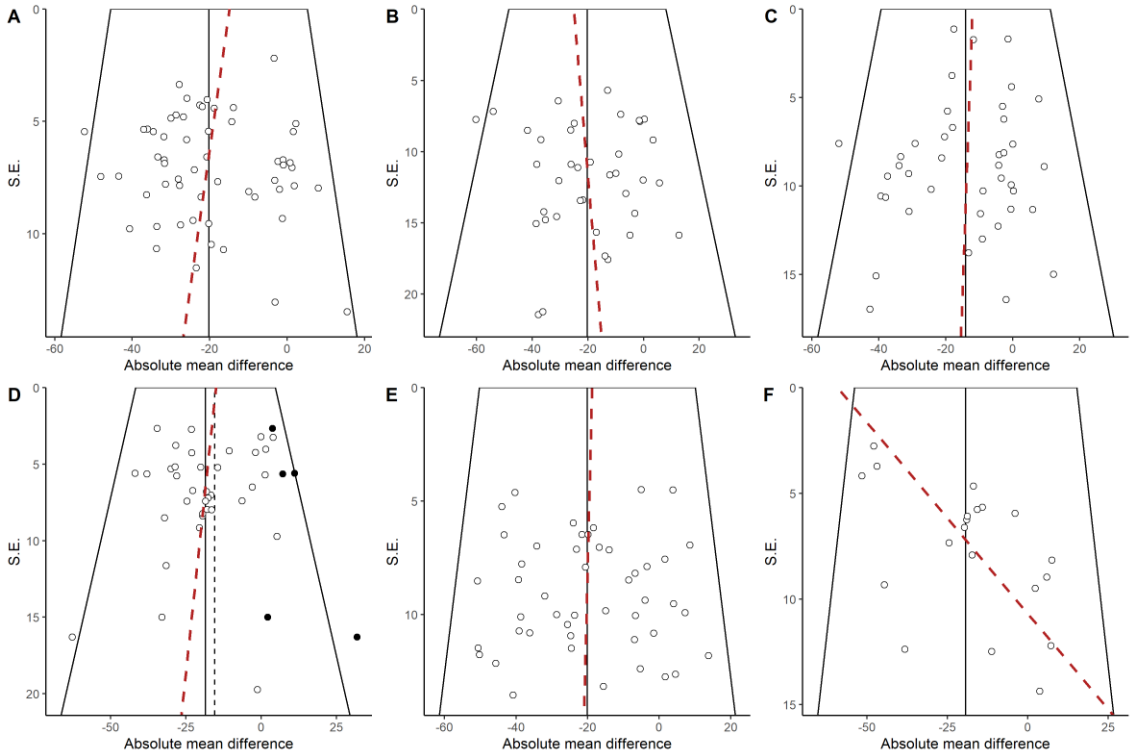

Figure 2 – Funnel plots and Egger's regression. Each observed study is represented by the white circles and the red dashed line represents Egger's regression line. Black circles represent missing studies imputed by the trim-and-fill L_0 method. **(A)** Training, systemic. Egger's regression p-value = 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, Ioannidis and Trikalinos' 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%	$2.9x10^{-282}$		
Complete, intervention session and injection site as moderators	81.9%	$2.4x10^{-195}$	20.9%	$7.6x10^{-14}$
			7.4%	$9.0x10^{-5}$

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 (Moulin & Amaral, 2020) 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 reasonably large 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**.

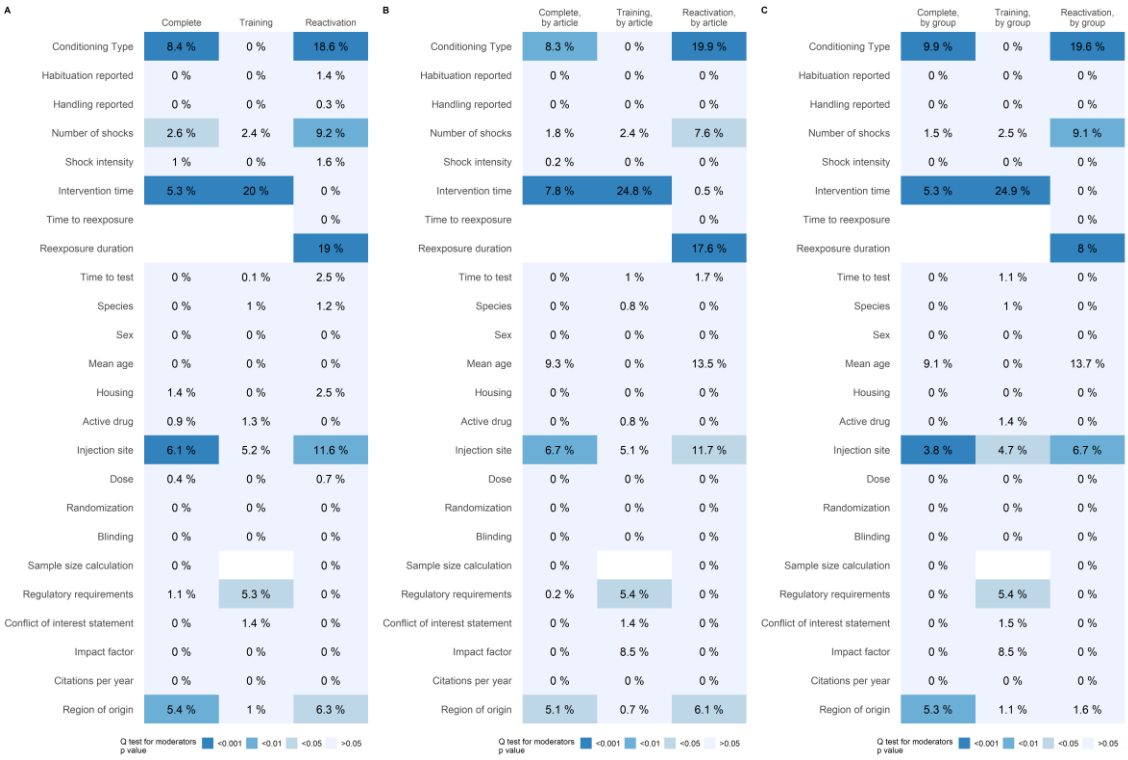

Figure 3 – Summary of univariable meta-regression models. Each column represents a dataset, while lines correspond to each variable. Cell colors represent p-values for Q tests of the respective variable as a moderator and numbers within cells are the R^2 value for the moderator in a univariable model. 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. **(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 due to 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.

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. Fig. 12**).

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 markedly 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.

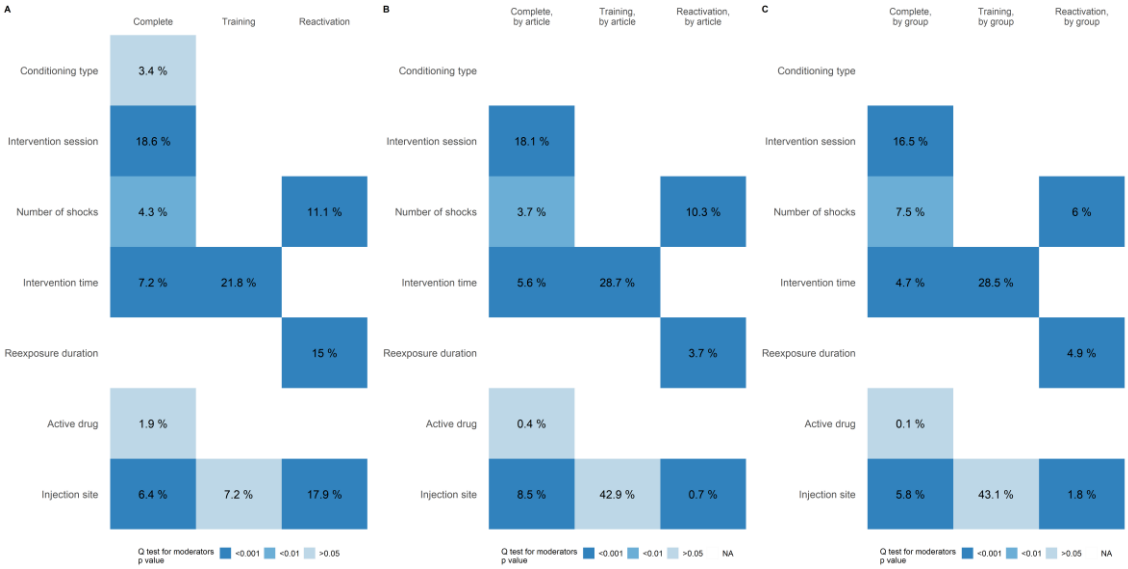

Figure 4 – Summary of multivariable meta-regression models. Each column represents a dataset and lines are variables selected in the best model for each scenario. Cell colors represent the p value for the Q test for each moderator in the meta-regression and numbers within are the decomposed R^2 value 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 a well-established finding in the field of behavioral neuroscience – the effect of protein synthesis inhibitors on fear conditioning. Beyond the results themselves, this effort also leads to some insights on how to best apply existing meta-analysis methods to discovery science, which presents a series of challenges to data synthesis.

One particular focus of our study was to investigate different ways to aggregate studies for data synthesis. We initially combined studies by intervention session and by injection site, performing individual meta-analyses for each combination. This showed a very robust effect for interventions in all three most studied sites (systemic injections, hippocampus and amygdala) on training and reconsolidation. This is unsurprising, as the requirement of protein synthesis for memory consolidation has been one of the dogmas in the field for decades (Davis & Squire, 1984). More recently, protein synthesis inhibitors also became the paradigmatic intervention to demonstrate the existence of a reconsolidation process (Nader, 2003; Nader et al., 2000)**.**

On the other hand, the effect of protein synthesis inhibition on fear extinction was not as robust as we expected, in part because far fewer studies were available. Evidence for the role of protein synthesis in extinction started to emerge in the early 2000s, though initially not in fear conditioning (Berman & Dudai, 2001; Vianna et al., 2001). Nevertheless, findings were not consensual from the start (Lattal & Abel, 2001). Moreover, when compared to reconsolidation, much less evidence has been accumulated since then. Because of this, our meta-analysis found a significant effect only for systemic interventions. That said, it's worth noting that most null effects on extinction from amygdala interventions come from a single study (Duvarci et al., 2006),

whereas the hippocampus analysis is heavily influenced by a study which, although nominally investigating extinction according to the authors, would be considered a reconsolidation experiment by most standards (Fischer et al., 2004).

Even though the training and reconsolidation data are robust on the aggregate, there is still a large amount of heterogeneity among individual experiments for virtually all meta-analyses performed, as found in previous work (Schroyens et al., 2021). When articles were included as an additional level, most of this heterogeneity was observed between different experiments within studies, rather than between studies. This may reflect the fact that behavioral neuroscience articles typically include many experiments varying protocol elements as tests of boundary conditions or negative controls. This presents a challenge for meta-analyses in discovery science, as inclusion criteria are rarely specific to the main results of a paper, and will commonly include many of these other experiments as well. The sheer volume of data in the training and reconsolidation meta-analyses makes these negative results insufficient to call the existence of an effect into question – nevertheless, it is likely that they may bias effect size estimates downwards.

Interestingly, aggregating studies from all injection sites or from interventions in both training and reactivation sessions did not cause much of an increase in heterogeneity, perhaps because it was already large to start with. As the increased sample sizes obtained by combining datasets allowed greater power for exploring possible sources of variation and identifying potential moderators, we chose to use this approach for our meta-regressions.

Some of the included moderators can function as a sanity check of whether meta-regression is able to capture consensual knowledge in the literature. It is widely accepted, for instance, that the sensitivity of fear conditioning to protein synthesis inhibitors or other drugs is time-dependent, occurring up to a few hours after the initial learning of an association (Bourtchouladze et al., 1998; Tiunova et al., 1998). Accordingly, in the training dataset, we found the time between training session and drug administration to be a significant moderator in all univariable and multivariable models tested.

Similarly, it is well accepted that the occurrence of reconsolidation or extinction after memory reactivation depends on the degree of prediction error between the information presented during training session and the reactivation, which is classically modulated by varying the number of CS presentations (Pedreira et al., 2004; Sevenster et al., 2014) or the duration of contextual reexposure (Alfei et al., 2015; Suzuki et al., 2004). In agreement with this, our measure of reexposure duration, which incorporated both of these manipulations, was found to be a significant moderator in the reactivation dataset. Interestingly, however, it had a much smaller effect size when we considered the research group as an additional level in the analysis.

Other postulated boundary conditions for reconsolidation, however, were not observed as clearly in our meta-analysis. Many studies indicate training strength as a moderator that can determine the occurrence of reconsolidation, extinction or neither (Eisenberg et al., 2003; Suzuki et al., 2004; Wang et al., 2009). This relation, however, is not linear: while very low training strength has been shown to favor extinction rather than reconsolidation (Eisenberg et al., 2003), very strong memories have been suggested to be more resistant to destabilization (Wang et al., 2009). Our meta-regressions found that a higher number of shocks led to a greater amnestic effect of protein synthesis inhibitors; that said, this effect was largely driven by a few

experiments with a very large number of shocks: when these were removed, the opposite association was observed. In contrast, shock intensity did not appear to be a relevant moderator.

Another boundary condition that has been repeatedly mentioned in the literature is memory age (i.e. time between training and reexposure). This was initially proposed by Milekic & Alberini (2002), but was followed by studies that either contradicted this finding (Debiec et al., 2002) or found it to be dependent on reexposure duration (Suzuki et al., 2004). Although our metanalysis included many of these studies, we did not find any influence of memory age on the effect of protein synthesis inhibitors during reactivation. Nevertheless, it is worth pointing out that our model does not include interactions between moderators; thus, if memory age has an effect that is modulated by other factors, it could have been obscured by confounding. It is also possible that this effect might be limited to very remote memories: studies with more than 20 days between training and reexposure did seem to have smaller effects in our sample, but this was offset in the analysis by large effects with reexposures occurring between 1 and 20 days after training.

Interestingly, type of conditioning (e.g. tone vs. contextual) was consistently detected as a significant moderator in our univariable models, an effect that was mostly driven by a high prevalence of near-null results in contextual conditioning experiments. Although various studies have evaluated the differences between the consolidation of contextual and cued fear memories (Kochli et al., 2015; Parsons et al., 2006; Stiedl et al., 1999), direct comparisons of the sensitivity of both protocols to reconsolidation have been lacking, largely because they tend to be used by different research groups. Importantly, however, the effect we found was nearly absent from multivariable models, perhaps due to its interaction with other variables such as the site of injection, as both types of conditioning tend to be studied in relation to different brain structures.

The changes in the relative importance of different moderators when multivariable and/or multilevel models are used underscore the importance to use these kinds of methods to try to minimize confounding. Although we cannot say for certain whether models become more accurate as they grow more complex, it’s worth noting that the effect of intervention time, perhaps the best-established moderator among those we studied, became stronger when multiple levels were added. On the other hand, factors such as the region of origin or the presence of a conflict of interest statement, which seem likely to be spurious, disappear when multilevel models are used and research group is taken into account. This suggests that meta-regressions that do not take these precautions in the literature might be at risk of detecting spurious moderators due to confounding.

Even after moderators and additional levels are included in the models, much heterogeneity remains, suggesting that classic boundary conditions alone are not sufficient to explain discrepancies in the literature. This is made clear by recent failures in trying to replicate post-retrieval amnesia findings from published studies, despite the use of similar protocols (Luyten et al., 2021; Rotondo et al., 2022). It could be that minor differences in reactivation protocols, factors such as anxiety (Soeter & Kindt, 2013), subtleties in animal strains and housing conditions (Schimanski & Nguyen, 2004) or interactions between moderators, all of which would not have been captured by our meta-analysis, can explain this additional heterogeneity.

Still, one should not lose sight of the possibility that heterogeneity and replication failures are not the result of methodological variation, but rather of biases in

the primary literature. In this sense, it's worth pointing out that the prevalence of randomization and sample size calculations in our sample was low, in accordance with previous surveys of preclinical studies (Macleod et al., 2015). Encouragingly, however, fear conditioning seems to have a much higher level of blinded outcome assessment than other areas of animal research, mostly due to the availability of software for automated freezing detection. Another interesting finding is the near absence of studies using female animals, which has also been pointed out by previous studies (Carneiro et al., 2018) and indicates striking lack of compliance with current recommendations from various agencies (Clayton & Collins, 2014; Tannenbaum et al., 2019).

Another simple explanation for apparent divergences in the literature is random error, which can lead to effect size vibration across statistical thresholds, particularly when power is low. Using the respective meta-analytical effect size estimates as references, we found the average power of individual studies in our sample to be 63%, which is in line with a previous estimate for the fear conditioning literature in general (Carneiro et al., 2018). Although this is higher than estimates for other fields of neuroscience (Button et al., 2013; Nord et al., 2017), it still means that more than one third of studies should fail to find statistically significant differences between groups in experiments with average effects. This suggests that lack of replication between results based on statistical significance criteria is expected to occur quite frequently.

Interestingly, however, we found little evidence of publication bias in our analyses. This is in stark contrast to what was found by Schroyens et al (2021), who showed clear evidence of publication bias when analyzing the effects of anisomycin, propranolol, MK-801 and midazolam on contextual fear reconsolidation. Interestingly, the evidence for publication bias in their study seems to be less clear for anisomycin than for the other drugs, suggesting that there might be less bias in the literature on protein synthesis inhibitors. Nevertheless, one should consider that our ability to detect publication bias may have been limited by the inclusion of experiments testing boundary conditions or negative controls, which are expected to yield negative results. In these cases, publication bias would be expected to act on the opposite direction (i.e. suppressing statistically significant results), thus complicating its detection on the sample as a whole. It is also important to note that publication bias assessments can be limited by the presence of true heterogeneity between studies (Duval, 2005), which is quite large in our sample.

In conclusion, we have found that the use of meta-analysis and meta-regression was able to capture some well-established findings in the fear conditioning literature. This includes both the robustness of the effect of protein synthesis inhibitors on consolidation and reconsolidation and its dependence on paradigmatic moderators such as the timing of post-training interventions and the amount of reexposure (as measured by duration or number of shocks) in reactivation experiments.

That said, not every boundary condition typically mentioned in the literature was detected: variables such as memory age, for example, had nearly null effects on reconsolidation sensitivity. This may indicate that the primary literature on these topics is more controversial than it is usually taken to be, or alternatively that meta-regression methods might be failing to capture real effects due to noise, heterogeneity, non-linearity of effects or interactions between variables. This leads to an interesting conundrum – when meta-regression doesn't agree with data from individual studies that have tested a particular moderator, where should we place our trust?

Unfortunately, there is no easy answer to this question. On the one hand, individual studies are typically underpowered, subject to various types of bias and should be taken with a grain of salt. On the other, meta-regression findings are observational in nature, and thus inevitably prone to confounding. The use of multilevel models and multivariable regression can help to mitigate this problem but cannot fully eliminate it. Potential moderators stemming from meta-regression should thus be considered tentative hypotheses that should ideally be tested in empirical studies designed for this purpose Although data synthesis can be useful for evaluating the consistency of the primary literature, it incorporates most of its biases, and thus cannot fully substitute for confirmatory, high-powered replication studies (Amaral & Neves, 2021).

A final advantage of empirically testing findings from meta-analysis is to formally compare the predictions of different types of models in terms of their accuracy. Although we argue in favor of using multilevel modelling and multivariable regression from a conceptual perspective, 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, albeit with a few important differences.

All of these issues, however, will only be resolved when systematic iteration between data synthesis and experimentation becomes more common in discovery science. This requires engaging researchers more widely in systematic reviews, as well as making these more efficient through automated tools (Bannach-Brown et al., 2021) and more structured primary data. It also involves building up infrastructure for confirmatory research that can go beyond the work of individual labs and validate meta-analytical findings empirically (Amaral & Neves, 2021). Ultimately, such interplay may produce a virtuous cycle that can improve the robustness of basic science theories 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.

Funding

This study was supported by grants E-26/111.277/2014, E-26/201.544/2014 and E-26/203.222/2017 from FAPERJ.

C.F.D.C. received scholarships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and an award from Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) for the conduction of this project (Edital bolsa TEC Mestrado Nota 10 2017).

Contributor roles (CRediT taxonomy - https://credit.niso.org/)

C.F.D.C.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing.

F.E.A.: Investigation, Writing – original draft, Writing – review & editing.

O.B.A.: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

References

Abate, G., Colazingari, S., Accoto, A., Conversi, D., & Bevilacqua, A. (2018). Dendritic spine density and EphrinB2 levels of hippocampal and anterior cingulate cortex neurons increase sequentially during formation of recent and remote fear memory in the mouse. *Behavioural Brain Research*, *344*, 120–131.

Abel, T., Nguyen, P. V., Barad, M., Deuel, T. A. S., Kandel, E. R., & Bourtchouladze, R. (1997). Genetic demonstration of a role for PKA in the late phase of LTP and in hippocampus-based long-term memory. *Cell*, *88*(5), 615–626.

Ageta, H., Ikegami, S., Miura, M., Masuda, M., Migishima, R., Hino, T., Takashima, N., Murayama, A., Sugino, H., Setou, M., et al. (2010). Activin plays a key role in the maintenance of long-term memory and late-LTP. *Learning and Memory*, *17*(4), 176–185.

Alfei, J. M., Monti, R. I. F., Molina, V. A., Bueno, A. M., & Urcelay, G. P. (2015). Prediction error and trace dominance determine the fate of fear memories after post-training manipulations. *Learning and Memory*, *22*(8), 385–400.

Almeida-Corrêa, S., Moulin, T. C., Carneiro, C. F. D., Gonçalves, M. M. C., Junqueira, L. S., & Amaral, O. B. (2015). Calcineurin inhibition blocks within-, but not between-session fear extinction in mice. *Learning & Memory*, *22*(3), 159–169.

Amaral, O. B., & Neves, K. (2021). Reproducibility: Expect less of the scientific paper. *Nature,* 597(7876), 329–331.

Amaral, O. B., Neves, K., Wasilewska-Sampaio, A. P., & Carneiro, C. F. (2019). The brazilian reproducibility initiative. e*Life*, *8,* e41602.

Apergis-Schoute, A. M., Dębiec, J., Doyère, V., LeDoux, J. E., & Schafe, G. E. (2005). Auditory fear conditioning and long-term potentiation in the lateral amygdala require ERK/MAP kinase signaling in the auditory thalamus: A role for presynaptic plasticity in the fear system. *Journal of Neuroscience*, *25*(24), 5730–5739.

Awad, W., Ferreira, G., & Maroun, M. (2015). Dissociation of the Role of Infralimbic Cortex in Learning and Consolidation of Extinction of Recent and Remote Aversion Memory. *Neuropsychopharmacology*, *40*(11), 2566–2575.

Bal, N. V., Rysakova, M. P., Vinarskaya, A. K., Ivanova, V., Zuzina, A. B., & Balaban, P. M. (2017). Cued memory reconsolidation in rats requires nitric oxide. *European Journal of Neuroscience*, *45*(5), 643–647.

Ballarini, F., Moncada, D., Martinez, M. C., Alen, N., & Viola, H. (2009). Behavioral tagging is a general mechanism of long-term memory formation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(34), 14599–14604.

Bannach-Brown, A., Hair, K., Bahor, Z., Soliman, N., MacLeod, M., & Liao, J. (2021). Technological advances in preclinical meta-research. *BMJ Open Science,* 5(1), e100131.

Barnes, P., & Thomas, K. L. (2008). Proteolysis of proBDNF Is a Key Regulator in the Formation of Memory. *PLoS ONE*, *3*(9), e3248.

Barrientos, R. M., Higgins, E. A., Sprunger, D. B., Watkins, L. R., Rudy, J. W., & Maier, S. F. (2002). Memory for context is impaired by injecting anisomycin into dorsal hippocampus following context exploration. *Behavioural Brain Research*, *134*(1–2), 299–306.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533.

Bekinschtein, P., Cammarota, M., Igaz, L. M., Bevilaqua, L. R. M., Izquierdo, I., & Medina, J. H. (2007). Persistence of Long-Term Memory Storage Requires a Late Protein Synthesis- and BDNF- Dependent Phase in the Hippocampus. *Neuron*, *53*(2), 261–277.

Berman, D. E., & Dudai, Y. (2001). Memory extinction, learning anew, and learning the new: Dissociations in the molecular machinery of learning in cortex. *Science*, *291*(5512), 2417–2419.

Bhattacharya, S., Kimble, W., Buabeid, M., Bhattacharya, D., Bloemer, J., Alhowail, A., Reed, M., Dhanasekaran, M., Escobar, M., & Suppiramaniam, V. (2017). Altered AMPA receptor expression plays an important role in inducing bidirectional synaptic plasticity during contextual fear memory reconsolidation. *Neurobiology of Learning and Memory*, *139*, 98–108.

Biedenkapp, J. C., & Rudy, J. W. (2004). Context memories and reactivation: Constraints on the reconsolidation hypothesis. *Behavioral Neuroscience*, *118*(5), 956–964.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Blum, S., Runyan, J. D., & Dash, P. K. (2006). Inhibition of prefrontal protein synthesis following recall does not disrupt memory for trace fear conditioning. *BMC Neuroscience*, *7*, 67.

Blundell, J., Kouser, M., & Powell, C. M. (2008). Systemic inhibition of mammalian target of rapamycin inhibits fear memory reconsolidation. *Neurobiology of Learning and Memory*, *90*(1), 28–35.

Bourtchouladze, R., Abel, T., Berman, N., Gordon, R., Lapidus, K., & Kandel, E. R. (1998). Different training procedures recruit either one or two critical periods for contextual memory consolidation, each of which requires protein synthesis and PKA. *Learning and Memory*, *5*(4–5), 365–374.

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

Cai, W. H., Blundell, J., Han, J., Greene, R. W., & Powell, C. M. (2006). Postreactivation glucocorticoids impair recall of established fear memory. *Journal of Neuroscience*, *26*(37), 9560–9566.

Carneiro, C. F. D., Moulin, T. C., Macleod, M. R., & Amaral, O. B. (2018). Effect size and statistical power in the rodent fear conditioning literature—A systematic review. *PLoS ONE*, 13(4), e0196258.

Chai, N., Liu, J.-F., Xue, Y.-X., Yang, C., Yan, W., Wang, H.-M., Luo, Y.-X., Shi, H.-S., Wang, J.-S., Bao, Y.-P., et al. (2014). Delayed Noradrenergic Activation in the Dorsal Hippocampus Promotes the Long-Term Persistence of Extinguished Fear. *Neuropsychopharmacology*, *39*(8), 1933–1945.

Clayton, J. A., & Collins, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature*, *509*(7500), 282–283.

Davis, H. P., & Squire, L. R. (1984). Protein synthesis and memory: A review. *Psychological Bulletin*, *96*(3), 518–559.

de Carvalho Myskiw, J., Furini, C. R. G., Schmidt, B., Ferreira, F., & Izquierdo, I. (2015). Extinction learning, which consists of the inhibition of retrieval, can be learned without retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(2), E230-3.

Debiec, J., Díaz-Mataix, L., Bush, D. E. A., Doyère, V., & Ledoux, J. E. (2010). The amygdala encodes specific sensory features of an aversive reinforcer. *Nature Neuroscience*, *13*(5), 536–537.

Debiec, J., Doyère, V., Nader, K., & LeDoux, J. E. (2006). Directly reactivated, but not indirectly reactivated, memories undergo reconsolidation in the amygdala. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(9), 3428–3433.

Debiec, J., LeDoux, J. E., & Nader, K. (2002). Cellular and systems reconsolidation in the hippocampus. *Neuron*, *36*(3), 527–538.

Duval, S. (2005). The Trim and Fill Method. In H. Rothstein, A. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis: Prevention, assessment and adjustments*. John Wiley & Sons Ltd.

Duval, S., & Tweedie, R. (2000a). A Nonparametric “Trim and Fill” Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, *95*(449), 89–98.

Duval, S., & Tweedie, R. (2000b). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463.

Duvarci, S., Mamou, C. ben, & Nader, K. (2006). Extinction is not a sufficient condition to prevent fear memories from undergoing reconsolidation in the basolateral amygdala. *European Journal of Neuroscience*, *24*(1), 249–260.

Duvarci, S., & Nader, K. (2004). Characterization of fear memory reconsolidation. *Journal of Neuroscience*, *24*(42), 9269–9275.

Duvarci, S., Nader, K., & LeDoux, J. E. (2005). Activation of extracellular signal-regulated kinase-mitogen-activated protein kinase cascade in the amygdala is required for memory reconsolidation of auditory fear conditioning. *European Journal of Neuroscience*, *21*(1), 283–289.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634.

Einarsson, E. Ö., & Nader, K. (2012). Involvement of the anterior cingulate cortex in formation, consolidation, and reconsolidation of recent and remote contextual fear memory. *Learning and Memory*, *19*(10), 449–452.

Eisenberg, M., Kobilo, T., Berman, D. E., & Dudai, Y. (2003). Stability of retrieved memory: Inverse correlation with trace dominance. *Science*, *301*(5636), 1102–1104.

Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *eLife*, *10,* e71601.

Ferrara, N. C., Cullen, P. K., Pullins, S. P., Rotondo, E. K., & Helmstetter, F. J. (2017). Input from the medial geniculate nucleus modulates amygdala encoding of fear memory discrimination. *Learning and Memory*, *24*(9), 414–421.

Fischer, A., Sananbenesi, F., Schrick, C., Spiess, J., & Radulovic, J. (2004). Distinct Roles of Hippocampal De Novo Protein Synthesis and Actin Rearrangement in Extinction of Contextual Fear. *Journal of Neuroscience*, *24*(8), 1962–1966.

França, T. F., & Monserrat, J. M. (2018). Reproducibility crisis in science or unrealistic expectations? *EMBO Reports*, *19*(6), e46008.

Frankland, P. W., Ding, H. K., Takahashi, E., Suzuki, A., Kida, S., & Silva, A. J. (2006). Stability of recent and remote contextual fear memory. *Learning and Memory*, *13*(4), 451–457.

Frankland, P. W., Josselyn, S. A., Anagnostaras, S. G., Kogan, J. H., Takahashi, E., & Silva, A. J. (2004). Consolidation of CS and US representations in associative fear conditioning. *Hippocampus*, *14*(5), 557–569.

Girardi, B. A., Ribeiro, D. A., Signor, C., Muller, M., Gais, M. A., Mello, C. F., & Rubin, M. A. (2016). Spermidine-induced improvement of reconsolidation of memory involves calcium-dependent protein kinase in rats. *Learning & Memory*, *23*(1), 21–28.

Gould, T. J., Wilkinson, D. S., Yildirim, E., Poole, R. L. F., Leach, P. T., & Simmons, S. J. (2014). Nicotine shifts the temporal activation of hippocampal protein kinase A and extracellular signal-regulated kinase 1/2 to enhance long-term, but not short-term, hippocampus-dependent memory. *Neurobiology of Learning and Memory*, *109*, 151–159.

Haubrich, J., Crestani, A. P., Cassini, L. F., Santana, F., Sierra, R. O., Alvares, L. de O., & Quillfeldt, J. A. (2015). Reconsolidation Allows Fear Memory to Be Updated to a Less Aversive Level through the Incorporation of Appetitive Information. *Neuropsychopharmacology*, *40*(2), 315–326.

Haubrich, J., Machado, A., Boos, F. Z., Crestani, A. P., Sierra, R. O., Alvares, L. D. O., & Quillfeldt, J. A. (2017). Enhancement of extinction memory by pharmacological and behavioral interventions targeted to its reactivation. *Scientific Reports*, *7*(1), 1–11.

Hoeffer, C. A., Cowansage, K. K., Arnold, E. C., Banko, J. L., Moerke, N. J., Rodriguez, R., Schmidt, E. K., Klosi, E., Chorev, M., Lloyd, R. E., et al. (2011). Inhibition of the interactions between eukaryotic initiation factors 4E and 4G impairs long-term associative memory consolidation but not reconsolidation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(8), 3383–3388.

Holehonnur, R., Phensy, A. J., Kim, L. J., Milivojevic, M., Vuong, D., Daison, D. K., Alex, S., Tiner, M., Jones, L. E., Kroener, S., & Ploski, J. E. (2016). Increasing the GluN2A/GluN2B ratio in neurons of the mouse basal and lateral amygdala inhibits the modification of an existing fear memory trace. *Journal of Neuroscience*, *36*(36), 9490–9504.

Huff, N. C., & Rudy, J. W. (2004). The amygdala modulates hippocampus-dependent context memory formation and stores cue-shock associations. *Behavioral Neuroscience*, *118*(1), 53–62.

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*(3), 245–253.

Jarome, T. J., Ferrara, N. C., Kwapis, J. L., & Helmstetter, F. J. (2015). Contextual Information Drives the Reconsolidation-Dependent Updating of Retrieved Fear Memories. *Neuropsychopharmacology, 40*(13), 3044–3052.

Jarome, T. J., Ferrara, N. C., Kwapis, J. L., & Helmstetter, F. J. (2016). CaMKII regulates proteasome phosphorylation and activity and promotes memory destabilization following retrieval. *Neurobiology of Learning and Memory*, *128*, 103–109.

Jarome, T. J., Kwapis, J. L., Werner, C. T., Parsons, R. G., Gafford, G. M., & Helmstetter, F. J. (2012). The timing of multiple retrieval events can alter GluR1 phosphorylation and the requirement for protein synthesis in fear memory reconsolidation. *Learning and Memory*, *19*(7), 300–306.

Jin, X., Qi, X., Yang, X., Li, B. (2007). Protein synthesis inhibition in the basolateral nucleus of amygdala facilitates extinction of auditory fear memory. *Chinese Science Bulletin,* 52, 2532–2542.

Kim, J., Song, B., Hong, I., Kim, J., Lee, J., Park, S., Yong Eom, J., Lee, C. J., Lee, S., & Choi, S. (2010). Reactivation of Fear Memory Renders Consolidated Amygdala Synapses Labile. *Journal of Neuroscience*, *30*(28), 9631–9640.

Kishioka, A., Uemura, T., Fukushima, F., & Mishina, M. (2013). Consolidation of auditory fear memories formed by weak unconditioned stimuli requires NMDA receptor activation and de novo protein synthesis in the striatum. *Molecular Brain*, *6*, 17.

Kochli, D. E., Campbell, T. L., Hollingsworth, E. W., Lab, R. S., Postle, A. F., Perry, M. M., Mordzinski, V. M., & Quinn, J. J. (2018). Combined administration of MK-801 and cycloheximide produces a delayed potentiation of fear discrimination memory extinction. *Behavioral Neuroscience*, *132*(2), 99–105.

Kochli, D. E., Thompson, E. C., Fricke, E. A., Postle, A. F., & Quinn, J. J. (2015). The amygdala is critical for trace, delay, and contextual fear conditioning. *Learning and Memory*, *22*(2), 92–100.

Kumar, T., & Jha, S. K. (2017). Influence of cued-fear conditioning and its impairment on NREM sleep. *Neurobiology of Learning and Memory*, *144*, 155–165.

Kwak, C., Choi, J.-H., Bakes, J. T., Lee, K., & Kaang, B.-K. (2012). Effect of intensity of unconditional stimulus on reconsolidation of contextual fear memory. *The Korean Journal of Physiology & Pharmacology*, *16*(5), 293–296.

Kwapis, J. L., Jarome, T. J., Ferrara, N. C., & Helmstetter, F. J. (2017). Updating Procedures can Reorganize the Neural Circuit Supporting a Fear Memory. *Neuropsychopharmacology*, 42, 1688-1697.

Kwapis, J. L., Jarome, T. J., Lee, J. L., & Helmstetter, F. J. (2015). The retrosplenial cortex is involved in the formation of memory for context and trace fear conditioning. *Neurobiology of Learning and Memory*, *123*, 110–116.

Kwapis, J. L., Jarome, T. J., Schiff, J. C., & Helmstetter, F. J. (2011). Memory consolidation in both trace and delay fear conditioning is disrupted by intra-amygdala infusion of the protein synthesis inhibitor anisomycin. *Learning & Memory* , *18*(11), 728–732.

Lattal, K. M., & Abel, T. (2001). Different requirements for protein synthesis in acquisition and extinction of spatial preferences and context-evoked fear. *Journal of Neuroscience*, *21*(15), 5773–5780.

Lattal, K. M., & Abel, T. (2004). Behavioral impairments caused by injections of the protein synthesis inhibitor anisomycin after contextual retrieval reverse with time. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4667–4672.

Lee, S. H., Choi, J. H., Lee, N., Lee, H. R., Kim, J. I., Yu, N. K., Choi, S. L., Lee, S. H., Kim, H., & Kaang, B. K. (2008). Synaptic protein degradation underlies destabilization of retrieved fear memory. *Science*, *319*(5867), 1253–1256.

Lopez, J., Gamache, K., Schneider, R., & Nader, K. (2015). Memory retrieval requires ongoing protein synthesis and NMDA receptor activity-mediated AMPA receptor trafficking. *Journal of Neuroscience*, *35*(6), 2465–2475.

Luyten, L., Schnell, A. E., Schroyens, N., & Beckers, T. (2021). Lack of drug-induced post-retrieval amnesia for auditory fear memories in rats. *BMC Biology*, *19*(1), 17.

Mac Callum, P. E., Hebert, M., Adamec, R. E., & Blundell, J. (2014). Systemic inhibition of mTOR kinase via rapamycin disrupts consolidation and reconsolidation of auditory fear memory. *Neurobiology of Learning and Memory*, *112*, 176–185.

Maciejak, P., Szyndler, J., Lehner, M., Turzyńska, D., Sobolewska, A., Bidziński, A., & Płaźnik, A. (2010). The differential effects of protein synthesis inhibition on the expression and reconsolidation of pentylenetetrazole kindled seizures. *Epilepsy and Behavior*, *18*(3), 193–200.

Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., Hirst, T., Hemblade, R., Bahor, Z., Nunes-Fonseca, C., et al. (2015). Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLOS Biology*, *13*(10), e1002273.

Mamiya, N., Fukushima, H., Suzuki, A., Matsuyama, Z., Homma, S., Frankland, P. W., & Kida, S. (2009). Brain region-specific gene expression activation required for reconsolidation and extinction of contextual fear memory. *Journal of Neuroscience*, *29*(2), 402–413.

Mamou, C. B., Gamache, K., & Nader, K. (2006). NMDA receptors are critical for unleashing consolidated auditory fear memories. *Nature Neuroscience*, *9*(10), 1237–1239.

Milekic, M. H., & Alberini, C. M. (2002). Temporally graded requirement for protein synthesis following memory reactivation. *Neuron*, *36*(3), 521–525.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., et al. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097.

Motanis, H., & Maroun, M. (2012). Differential involvement of protein synthesis and actin rearrangement in the reacquisition of contextual fear conditioning. *Hippocampus*, *22*(3), 494–500.

Moulin, T. C., & Amaral, O. B. (2020). Using collaboration networks to identify authorship dependence in meta-analysis results. *Research Synthesis Methods*, *11*(5), jrsm.1430.

Nader, K. (2003). Memory traces unbound. *Trends in Neurosciences*, *26*(2), 65–72.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726.

Nakayama, D., Baraki, Z., Onoue, K., Ikegaya, Y., Matsuki, N., & Nomura, H. (2015). Frontal association cortex is engaged in stimulus integration during associative learning. *Current Biology*, *25*(1), 117–123.

Nakayama, D., Yamasaki, Y., Matsuki, N., & Nomura, H. (2013). Post-retrieval late process contributes to persistence of reactivated fear memory. *Learning and Memory*, *20*(6), 307–310.

Nomura, H., & Matsuki, N. (2008). Ethanol enhances reactivated fear memories. *Neuropsychopharmacology*, *33*(12), 2912–2921.

Nord, C. L., Valton, V., Wood, J., & Roiser, J. P. (2017). Power-up: A reanalysis of 'power failure' in neuroscience using mixture modelling. *The Journal of Neuroscience*, *37*(34), 3592–16.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Parsons, R. G., Gafford, G. M., Baruch, D. E., Riedner, B. A., & Helmstetter, F. J. (2006). Long-term stability of fear memory depends on the synthesis of protein but not mRNA in the amygdala. *European Journal of Neuroscience*, *23*(7), 1853–1859.

Parsons, R. G., Riedner, B. A., Gafford, G. M., & Helmstetter, F. J. (2006). The formation of auditory fear memory requires the synthesis of protein and mRNA in the auditory thalamus. *Neuroscience*, *141*(3), 1163–1170.

Casanova, J., Madrid, C., Contreras, M., Rodriguez, M., Vasquez, M. & Torrealba, F., (2016). A role for the interoceptive insular cortex in the consolidation of learned fear. *Behavioural Brain Research*, *296*, 70–77.

Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning and Memory*, *11*(5), 579–585.

Poulos, A. M., Ponnusamy, R., Dong, H.-W., & Fanselow, M. S. (2010). Compensation in the neural circuitry of fear conditioning awakens learning circuits in the bed nuclei of the stria terminalis. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14881–14886.

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews. Drug Discovery*, *10*(9), 712.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Radwanska, K., Medvedev, N. I., Pereira, G. S., Engmann, O., Thiede, N., Moraes, M. F. D., Villers, A., Irvine, E. E., Maunganidze, N. S., Pyza, E. M., et al. (2011). Mechanism for long-term memory formation when synaptic strengthening is impaired. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(45), 18471–18475.

Reis, D. S., Jarome, T. J., & Helmstetter, F. J. (2013). Memory formation for trace fear conditioning requires ubiquitin-proteasome mediated protein degradation in the prefrontal cortex. *Frontiers in Behavioral Neuroscience*, *7*(OCT), 150.

Remaud, J., Ceccom, J., Carponcy, J., Dugué, L., Menchon, G., Pech, S., Halley, H., Francés, B., & Dahan, L. (2014). Anisomycin injection in area CA3 of the hippocampus impairs both short-term and long-term memories of contextual fear. *Learning and Memory*, *21*(6), 311–315.

Rizzo, V., Touzani, K., Raveendra, B. L., Swarnkar, S., Lora, J., Kadakkuzha, B. M., Liu, X. A., Zhang, C., Betel, D., Stackman, R. W., & Puthanveettil, S. V. (2017). Encoding of Contextual Fear Memory Requires De Novo Proteins in the Prelimbic Cortex. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(2), 158–169.

Rotondo, F., Biddle, K., Chen, J., Ferencik, J., d'Esneval, M., & Milton, A. L. (2022). Lack of Effect of Propranolol on the Reconsolidation of Conditioned Fear Memory due to a Failure to Engage Memory Destabilisation. *Neuroscience*, *480*, 9–18.

Roy, D. S., Muralidhar, S., Smith, L. M., & Tonegawa, S. (2017). Silent memory engrams as the basis for retrograde amnesia. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(46), E9972–E9979.

Rudy, J. W., & Matus-Amat, P. (2005). The ventral hippocampus supports a memory representation of context and contextual fear conditioning: Implications for a unitary function of the hippocampus. *Behavioral Neuroscience*, *119*(1), 154–163.

Runyan, J. D., & Dash, P. K. (2005). Inhibition of hippocampal protein synthesis following recall disrupts expression of episodic-like memory in trace conditioning. *Hippocampus*, *15*(3), 333–339.

Sacchetti, B., Sacco, T., & Strata, P. (2007). Reversible inactivation of amygdala and cerebellum but not perirhinal cortex impairs reactivated fear memories. *European Journal of Neuroscience*, *25*(9), 2875–2884.

Santini, E., Ge, H., Ren, K., Peña De Ortiz, S., & Quirk, G. J. (2004). Consolidation of fear extinction requires protein synthesis in the medial prefrontal cortex. *Journal of Neuroscience*, *24*(25), 5704–5710.

Schafe, G. E., & LeDoux, J. E. (2000). Memory consolidation of auditory pavlovian fear conditioning requires protein synthesis and protein kinase A in the amygdala. *The Journal of Neuroscience*, *20*(18), RC96–RC96.

Scharf, M. T., Woo, N. H., Lattal, K. M., Young, J. Z., Nguyen, P. V., & Abel, T. (2002). Protein Synthesis Is Required for the Enhancement of Long-Term Potentiation and Long-Term Memory by Spaced Training. *Journal of Neurophysiology*, *87*(6), 2770–2777.

Schimanski, L. A., & Nguyen, P. V. (2004). Multidisciplinary approaches for investigating the mechanisms of hippocampus-dependent memory: A focus on inbred mouse strains. *Neuroscience and Biobehavioral Reviews,* 28(5), 463–483.

Schmidt, S. D., Furini, C. R. G., Zinn, C. G., Cavalcante, L. E., Ferreira, F. F., Behling, J. A. K., Myskiw, J. C., & Izquierdo, I. (2017). Modulation of the consolidation and reconsolidation of fear memory by three different serotonin receptors in hippocampus. *Neurobiology of Learning and Memory*, *142*, 48–54.

Schroyens, N., Alfei, J. M., Schnell, A. E., Luyten, L., & Beckers, T. (2019). Limited replicability of drug-induced amnesia after contextual fear memory retrieval in rats. *Neurobiology of Learning and Memory*, *166*, 107105.

Schroyens, N., Sigwald, E. L., Van Den Noortgate, W., Beckers, T., & Luyten, L. (2021). Reactivation-dependent amnesia for contextual fear memories: Evidence for publication bias. *ENeuro*, *8*(1), 1–16.

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning and Memory*, *21*(11), 580–584.

Signor, C., Temp, F. R., Mello, C. F., Oliveira, M. S., Girardi, B. A., Gais, M. A., Funck, V. R., & Rubin, M. A. (2016). Intrahippocampal infusion of spermidine improves memory persistence: Involvement of protein kinase A. *Neurobiology of Learning and Memory*, *131*, 18–25.

Soeter, M., & Kindt, M. (2013). High Trait Anxiety: A Challenge for Disrupting Fear Memory Reconsolidation. *PLOS ONE*, *8*(11), e75239.

Song, Z., Chen, H., Xu, W., Wu, S., & Zhu, G. (2018). Basolateral amygdala calpain is required for extinction of contextual fear-memory. *Neurobiology of Learning and Memory*, *155*, 180–188.

Stafford, J. M., & Lattal, K. M. (2009). Direct comparisons of the size and persistence of anisomycin-induced consolidation and reconsolidation deficits. *Learning and Memory*, *16*(8), 494–503.

Stanley, D. J., & Spence, J. R. (2014). Expectations for Replications: Are Yours Realistic? *Perspectives on Psychological Science*, *9*(3), 305–318.

Steward, O., Popovich, P. G., Dietrich, W. D., & Kleitman, N. (2012). Replication and reproducibility in spinal cord injury research. *Experimental Neurology,* 233(2), 597–605.

Stiedl, O., Palve, M., Radulovic, J., Birkenfeld, K., & Spiess, J. (1999). Differential impairment of auditory and contextual fear conditioning by protein synthesis inhibition in C57BL/6N mice. *Behavioral Neuroscience*, *113*(3), 496–506.

Suzuki, A., Josselyn, S. A., Frankland, P. W., Masushige, S., Silva, A. J., & Kida, S. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *Journal of Neuroscience*, *24*(20), 4787–4795.

Suzuki, A., Mukawa, T., Tsukagoshi, A., Frankland, P. W., & Kida, S. (2008). Activation of LVGCCs and CB1 receptors required for destabilization of reactivated contextual fear memories. *Learning and Memory*, *15*(6), 426–433.

Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J., & Schiebinger, L. (2019). Sex and gender analysis improves science and engineering. *Nature*, *575*(7781), 137–146.

Tiunova, A. A., Anokhin, K. V., & Rose, S. P. R. (1998). Two critical periods of protein and glycoprotein synthesis in memory consolidation for visual categorization learning in chicks. *Learning and Memory*, *4*(5), 401–410.

Trent, S., Barnes, P., Hall, J., & Thomas, K. L. (2015). Rescue of long-term memory after reconsolidation blockade. *Nature Communications*, *6*, 1–7.

Vesterinen, H. M., Sena, E. S., Egan, K. J., Hirst, T. C., Churolov, L., Currie, G. L., Antonic, A., Howells, D. W., & Macleod, M. R. (2014). Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods*, *221*, 92–102.

Vianna, M. R., Szapiro, G., McGaugh, J. L., Medina, J. H., & Izquierdo, I. (2001). Retrieval of memory for fear-motivated training initiates extinction requiring protein synthesis in the rat hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12251–12254.

Viosca, J., De Armentia, M. L., Jancic, D., & Barco, A. (2009). Enhanced CREB-dependent gene expression increases the excitability of neurons in the basal amygdala and primes the consolidation of contextual and cued fear memory. *Learning and Memory*, *16*(3), 193–197.

von Hertzen, L. S. J. (2005). Memory Reconsolidation Engages Only a Subset of Immediate-Early Genes Induced during Consolidation. *Journal of Neuroscience*, *25*(8), 1935–1942.

Wang, S. H., De Oliveira Alvares, L., & Nader, K. (2009). Cellular and systems mechanisms of memory strength as a constraint on auditory fear reconsolidation. *Nature Neuroscience*, *12*(7), 905–912.

Wanisch, K., Tang, J., Mederer, A., & Wotjak, C. T. (2005). Trace fear conditioning depends on NMDA receptor activation and protein synthesis within the dorsal hippocampus of mice. *Behavioural Brain Research*, *157*(1), 63–69.

Wilensky, A. E., Schafe, G. E., Kristensen, M. P., & LeDoux, J. E. (2006). Rethinking the fear circuit: The central nucleus of the amygdala is required for the acquisition,

consolidation, and expression of pavlovian fear conditioning. *Journal of Neuroscience*, *26*(48), 12387–12396.

Wotjak, C. T. (2019). Sound check, stage design and screen plot – how to increase the comparability of fear conditioning and fear extinction experiments. *Psychopharmacology,* 236(1), 33–48.

Yang, C.-H., Huang, C.-C., & Hsu, K.-S. (2011). Generalization of fear inhibition by disrupting hippocampal protein synthesis-dependent reconsolidation process. *Neuropsychopharmacology*, *36*(10), 1992–2008.

Yim, A. J., Moraes, C. R. G., Ferreira, T. L., & Oliveira, M. G. M. (2006). Protein synthesis inhibition in the basolateral amygdala following retrieval does not impair expression of morphine-associated conditioned place preference. *Behavioural Brain Research*, *171*(1), 162–169.