MASTERARBEIT | MASTER'S THESIS

Titel | Title

The Impact of Acute Stress on Learning: A Meta-Analytic Review

verfasst von | submitted by

Lisa Steiner BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Claus Lamm

Abstract

Acute stress affects core cognitive processes, yet its impact on learning has been inconsistently reported. The present thesis provides a systematic and quantitative evaluation of this relationship through a three-level meta-analysis synthesizing findings from experimentally induced acute stress paradigms. A comprehensive literature search identified studies with adult participants exposed to acute stressors, and reporting learning-related outcomes. Thirty studies met the inclusion criteria, resulting in multiple extractable effect sizes which were then analysed using random-effects multilevel models accounting for dependencies within outcomes. Overall, acute stress showed a small to medium impairing effect on learning performance. Likelihood ratio tests indicated no within-study heterogeneity, whereas between-study heterogeneity was present. Moderator analyses revealed no differences between learning types, stressor types, or post-learning stress. In contrast, concurrent stress was less impairing than pre-learning stress, and within-subject designs showed significantly stronger negative effects than between-subject designs. Publication-bias diagnostics provided no evidence of small-study bias and consistently supported a negative underlying effect. The results demonstrate that acute stress reliably disrupts learning, particularly when stress precedes the learning episode and when within-person comparisons are possible. These findings help clarify previously inconsistent evidence and highlight the importance of methodological and contextual factors in stress–learning research.

Contents

Introduction

Acute stress plays a fundamental role in shaping human behaviour, cognition, and physiological responses. Understanding how stress alters core processes such as learning is therefore essential for gaining insight into adaptive and maladaptive mechanisms of human functioning. Against this background, the present work examines how acute stress affects learning and which factors influence this relationship.

Stress can be described as a threat to homeostasis that activates physiological and behavioural processes. The human body adapts to this homeostatic change through psychological mechanisms (emotional and cognitive changes), behavioural responses (such as the fight-or-flight reaction), and biological adjustments (alterations in autonomic and neuroendocrine function). These responses are controlled by two major effectors: the hypothalamic–pituitary–adrenal (HPA) axis and the sympathetic adreno-medullary (SAM) system (Bali & Jaggi, 2015). The endocrine response to stress involves activation of the HPA axis, culminating in the release of cortisol from the adrenal cortex. Activation of the SAM system triggers the release of the neurotransmitters catecholamines adrenaline and noradrenaline, which increase heart rate and blood pressure (Turner et al., 2020).

From a psychological perspective, stress can also be described as a state of unpleasant tension influenced by the quality, intensity, and duration of the stressor (Greif, 1991). Although research distinguishes negative stress (*distress*) from positive stress (*eustress*), the scientific literature largely focuses on the negative component (Nerdinger et al., 2011). Classic determinants of the stress response include novelty, unpredictability, and uncontrollability, factors identified as key triggers of physiological activation (Mason, 1968). A useful framework for conceptualizing the formation of stress is the Transactional Model of Stress (Lazarus & Folkman, 1984). This model views stress as an interactional and transactional process involving the cognitive and emotional appraisal of a situation and the person's perceived resources to cope with it. During the *primary appraisal* stage, individuals evaluate whether a situation poses a potential threat. If it is appraised as threatening, the *secondary appraisal* assesses whether available resources are sufficient to deal with the situation. Stress arises only if, after this re-evaluation, the individual concludes that their resources are not adequate to cope with the threat.

A crucial factor explaining why some people experience stress in certain situations and others do not is the availability of coping resources and strategies. Resources can be

understood as intrinsic, environmental, or material factors that support coping. Intrinsic resources, such as personality traits, strengthen resilience. Environmental resources include elements such as a strong social network, while material or monetary resources also function as important protective factors (Nerdinger et al., 2011).
Depending on the duration of the threatening situation, the stress response may be acute or chronic. Acute stress represents a temporary reaction characterized by symptoms such as anxiety or impaired concentration, lasting up to one month after exposure. Chronic stress, by contrast, leads to long-term or even permanent physical and psychological changes. Persistent activation of stress responses has been identified as a major risk factor for a wide variety of diseases (Bali & Jaggi, 2015).

For the assessment of stress in humans, various physiological and behavioural measures are commonly used, including cardiovascular indicators (e.g., blood pressure and heart rate), galvanic skin response (GSR), heart rate variability (HRV), pupil dilation, and muscle or skin sympathetic nerve activity (SSNA). Biochemical markers such as cortisol, salivary α-amylase, or norepinephrine are also frequently employed. Some studies additionally use self-report measures. To understand how these physiological and psychological responses are elicited under controlled conditions, numerous standardized laboratory stress paradigms have been developed. Some tasks focus mainly on the physical components of stress, while others include psychosocial elements. A widely used paradigm for inducing physical stress is the Cold Pressor Test (CPT). Participants immerse their hands in ice-cold water to induce physical pain (Lovallo, 1975). Due to its shorter duration and minimal personnel requirements, the CPT is economically advantageous compared to other methods. The CPT reliably activates the SAM axis but produces only minimal activation of the HPA axis. A variant of the CPT that strengthens HPA activation is the socially evaluated CPT (SECPT). In this paradigm, a social threat component is added by having an evaluator of the opposite sex present during the task. Even with this addition, cortisol responses remain relatively low, which may be explained by the absence of two major determinants of stress: unpredictability and uncontrollability.

A paradigm that reliably activates both the SAM and HPA axis is the Trier Social Stress Test (TSST) (Kirschbaum et al., 1993). To cause stress, participants are asked to deliver a speech, typically a simulated job interview, followed by a mental arithmetic task. These tasks alone do not consistently induce stress; therefore, the TSST includes a strong social-evaluative component. Both tasks must be performed in front of a panel of evaluators who are instructed not to provide positive feedback or social interaction (Bali & Jaggi, 2015).

Beyond the TSST, several other paradigms aim to combine different stress-inducing components to achieve reliable activation of stress systems. The Maastricht Acute Stress Test (MAST) combines the CPT and TSST to create a physiologically and psychologically challenging protocol (Smeets et al., 2012). Participants undergo five SECPT trials, and between these trials they complete socially evaluated arithmetic tasks analogous to the TSST. Another widely used protocol is the Montreal Imaging Stress Task (MIST) (Dedovic et al., 2005), which combines a stressful situation, such as the physical constraints and noise of MRI scanning, with computer-based arithmetic tasks. During the task, performance and time pressure are displayed, adding an additional stressful component. In addition to these tasks noise exposure is also used to induce stress in laboratory settings. Research shows that exposure to loud noise above 80 dB increases activity in both the SAM and HPA systems (Bali & Jaggi, 2015).

Understanding how these stress responses interact with cognitive processes requires a closer look at the mechanisms of human learning. Learning is a broad theoretical construct that can be described as a process leading to relatively permanent changes in an individual's behaviour or behavioural potential that arise from past experiences. Although learning mechanisms can only be inferred through performance, it is important to note that learning and performance are not equivalent. Performance may be impaired even after successful learning, which is essential to consider when analysing stress-induced impairments: decreased task performance may reflect impaired retrieval rather than unsuccessful learning.

A central distinction in learning mechanisms lies between automatic responses elicited by certain stimuli (Pavlovian or classical conditioning) and voluntary responses (operant or instrumental conditioning). In *Pavlovian conditioning*, repeated pairing of an unconditioned stimulus (US), which naturally provokes a response, with a neutral conditioned stimulus (CS), leads to a learned association. After learning, the response is elicited by the CS alone. The US may be appetitive (e.g., food) or aversive (e.g., electric shocks). This learning is involuntary and automatic.

In contrast, *instrumental* and *reinforcement learning* involve voluntary responses. In instrumental conditioning, individuals learn associations between their behaviour and its consequences. These consequences can take the form of positive reinforcement (presentation of a pleasant stimulus), negative reinforcement (removal of an aversive stimulus), or punishment (Gross, 2020). The Reinforcement Learning Theory (RLT) proposes that learning is driven by discrepancies between expected and actual outcomes, so-called prediction errors. When an outcome differs from what was anticipated, expectations must be substantially updated. If an outcome matches expectations, there is little new information, and learning is minimal. In research, reinforcement learning is often examined using tasks that require participants to learn associations between actions and outcomes (reward or punishment). These tasks are frequently probabilistic, meaning that rewards or punishments do not occur on every trial, requiring participants to learn the most frequent associations (Cutler et al., 2025).

After outlining how reinforcement learning relies on the comparison between expected and actual outcomes, it is important to note that these processes depend not only on action–outcome associations but also on the broader memory systems that store, update, and retrieve learned information. In this sense, reinforcement learning and memory are closely intertwined: without the ability to maintain and recall relevant information, neither stable learning nor flexible adaptation would be possible. Memory refers to the ability to store and retrieve learned information, knowledge, or behaviour and is typically divided into declarative and non-declarative systems. Classical conditioning and procedural memory fall under non-declarative processes, while declarative memory encompasses short-term memory, long-term memory, and working memory (Bear et al., 2018). Experimental tasks used to assess declarative memory include item-recognition tasks, such as paired-word learning (Kamp et al., 2019), and visual search tasks in which participants locate targets among distractors with embedded memory cues (Goldfarb et al., 2017).

Taken together, the wide range of learning mechanisms and memory processes provides a rich framework for understanding how stress might alter cognitive functioning, yet recent research paints an uneven picture of how acute stress shapes learning. Stress during memory retrieval impairs episodic discrimination processes (Nitschke et al., 2020) and hinders the recall of previously learned material more broadly (Schwabe, 2017). Evidence also indicates that acute psychosocial stress disrupts episodic memory processing in a stage-specific manner, with different effects emerging during encoding versus retrieval (Schwabe et al., 2012). At the same time, stress does not uniformly impair encoding: emotionally salient material can benefit from stress-related arousal, leading to enhanced episodic encoding under certain conditions (Ulrich-Lai & Herman, 2009). Beyond memory, acute stress has been shown to shift instrumental behaviour away from flexible, goal-directed control toward habitual responding and perseveration when contingencies change (Schwabe & Wolf, 2009). Findings from reward-based learning tasks reinforce this pattern, suggesting that acute stress impairs reward-seeking behaviour (Berghorst et al., 2013; Bogdan et al., 2011) and directly disrupts reward learning mechanisms (Carvalheiro et al., 2021). Taken together, these findings suggest that acute stress does not exert a uniform effect on cognition but interacts with the underlying learning system - often weakening flexible, model-based processes while biasing behaviour toward more rigid, habitual strategies.

The aim of this study is to systematically examine how acute stress affects learning outcomes across different experimental contexts. By synthesizing findings from a diverse body of research, the meta-analysis seeks to quantify the overall impact of acute stress on learning and to identify conditions under which this effect varies. In particular, the study investigates whether specific characteristics, such as stressor type, timing of stress induction, learning task, or sample-related factors, moderate the relationship between acute stress and learning. Through this approach, the study aims to clarify inconsistent findings in the literature and provide a more comprehensive understanding of how acute stress shapes learning processes.

Method

Literature Search

To identify relevant studies a systematic literature search was conducted in the electronic Databases Scopus, PubMed and Web of Science. Therefore, based on theoretical considerations, following search string was built:

("acute stress" OR "psychosocial stress" OR "cold pressor" OR "CPT" OR "TSST" OR "MAST" OR "MIST") AND ("reinforcement-learning" OR "reward learning" OR "punishment learning" OR "operant learning" OR "instrumental learning" OR "sequential decision making" OR "probabilistic learning" OR "bandit" OR "q-learning" OR "prediction error")

The literature search was conducted up to and including March 26, 2025. Accordingly, studies published up to that date were included. To ensure transparency, the search process was carefully documented in a PRISMA flowchart (Page et al., 2021).

Inclusion- and Exclusion Criteria

To be included in the analysis, studies had to meet the following criteria:

The inclusion and exclusion criteria for the meta-analytical study required that studies involved human participants who were adults (18+). They had to manipulate acute stress using methods such as the Cold Pressor Test (CPT), the Trier Social Stress Test (TSST), the Socially Evaluated Cold Pressor Test (SECPT), the Montreal Imaging Stress Task (MIST), the Maastricht Acute Stress Test (MAST), or variations thereof. Additionally, studies needed to include a non-stressed control group and report findings related to a learning task. Finally, sufficient data to compute effect sizes had to be reported in the article or obtained from the authors upon request.

Data Extraction

The following aspects of the included studies were coded: (i) Study characteristics: Name of the authors, year of publication, country of data collection, publication status (published/unpublished), and peer-reviewed status (yes/no). (ii) Study-specific details: Stress paradigms, tests used for learning assessment, learning type, stress time (pre-, postlearning or concurrent), and statistics for comparison. (iii) Sample description: Mean age of the experimental and control groups, sample size and number of participants per group, gender of the participants group allocation (randomized/non-randomized), and study design (between-subjects, within-subjects or mixed). (iv) Statistical parameters: Means and standard deviations, t values, beta coefficients, Chi-squared, Cohen's f, Cohen's d, Hedges'g and degrees of freedom were extracted if reported. Additionally, each study and effect size was assigned an individual identification number. This approach aligns with the three-level meta-analysis framework, which accounts for dependencies in the data arising from multiple effect sizes extracted from a single study (Assink & Wibbelink, 2016).

Statistical Analysis

All analyses were conducted in R using the metafor package (Viechtbauer, 2010). Multiple effect sizes per study were extracted since different learning outcomes were often investigated. However, traditional univariate meta-analytical approaches assume that effect sizes are independent. Because effect sizes from the same study are based on the same sample, they are likely to be more similar than effect sizes from different studies. This can lead to an overestimation of the overall effect, as studies contributing multiple effect sizes would disproportionately influence the mean effect compared to studies with only a single effect size. Several approaches exist to address this dependency. An approach that retains all available information while accounting for effect size dependencies is multilevel meta-analysis. Multilevel models incorporate all relevant effect sizes while considering dependency structures. The three-level meta-analysis estimates variance components at three levels: (1) sampling variance of effect sizes, (2) variance between effect sizes within the same study, and (3) variance between studies (Assink & Wibbelink, 2016).

The statistical analysis followed the guidelines for three-level models by Assink and Wibbelink (2016). Assuming that true effect sizes vary, an independent random-effects model was calculated for each of the four transfer types. The unique identification number assigned to each study ensured that effect sizes originating from the same study were recognized, allowing dependency to be accounted for in the analysis. First, the main effects were estimated to assess the strength of the relationships. To evaluate heterogeneity between studies (between-study variance) and within studies (within-study variance), two separate likelihood ratio tests were conducted for each transfer type. Since variance components can only deviate from zero in a positive direction, a one-sided test is recommended. As R reports *p*-values as two-sided by default, the *p*-values of the likelihood ratio tests were divided by two for interpretation. These tests provided insights into the significance of variance components. Since the null hypothesis states that one of the variance components is zero, a significant result indicates a statistically better model fit for the three-level model compared to a two-level model (Assink & Wibbelink, 2016).

Dealing with data from varying research designs

Because of the varying study designs and reported outcomes two independent meta-analyses were conducted. The primary analysis included studies or effect sizes that reflected clear group contrasts (stress vs. control group). To ensure comparability, effect sizes were standardized. Accordingly, Cohen's d was calculated and divided by the pooled standard deviation. Since Cohen's d tends to overestimate absolute effects in small samples, it was multiplied by the correction factor J and transformed into Hedges' g (Borenstein et al., 2021). Effect size interpretation followed established thresholds: small effect = 0.2, medium effect = 0.5, and large effect = 0.8 (Cohen, 1988).

The secondary model included studies that reflected interaction effects (variance analysis). The effect sizes were converted to Pearson's r and transformed to Fisher's z.

For better interpretability the results were converted back to r. Interpretation uses Cohen's thresholds: small effect = 0.1, medium effect = 0.3, and large effect = 0.5 (Cohen, 1988).

Direction of the Effect

In the primary as well as the secondary analysis higher scores indicate an impaired learning outcome. Higher effects of the control group compared to the stress group is coded as a negative Hedges' g and Fisher's z. Studies that used a different interpretation of the scores were analysed with an inverted effect size to fit the direction of the metanalysis.

Goal-directed versus Habitual Behaviour (explorative)

To assess whether acute stress shifts control from goal-directed to habitual responding an explorative meta-analysis restricted to studies employing outcome-devaluation, contingency-degradation, or closely related paradigms was conducted. Effect sizes were Fisher's z correlations (negative values indicate poorer goal-directed control in the stress group relative to control), calculated from reported r or converted from test statistics.

Heterogeneity

To determine whether the meta-analytic model should account for within-study variance a one-sided log-likelihood-ratio test was performed that compares the fit of the original model where the variance at level two and three is freely estimated to a model where the variance at level two is fixed to zero and only the variance at level three is freely estimated.

The significance of the between-study variance is tested in a similar way. The fit of the original model is compared to a model where the variance at level three is fixed to zero and only the two-level variance is freely estimated. For both tests the null hypothesis states that the within-study variance respectively the between-study variance equals zero and the alternative hypothesis states that the variance is greater than zero. A significant test indicates a better fit of the original three-level model compared to a two-level model.

Additionally, to the significance of the within-study and between-study variance the distribution of the total variance over the three levels was tested using the formulas by Cheung (2014). Because sampling-variance is based on sample size, and sample sizes vary across studies and effect sizes, Level-1 variance (sampling variance) cannot be treated as one fixed value, but it can be estimated using the aforementioned formula (Assink & Wibbelink, 2016).

Moderator Analyses

To address significant within-study and between-study variance, a subgroup analysis was conducted. Following variables were tested as possible moderators of the overall effect: (i) Learning type: The variable was coded into three theory-based categories. First, Instrumental and Reinforcement-Learning (RL) (e.g., probabilistic selection, reversal learning, outcome devaluation, feedback-based categorization). Second, Pavlovian (classical) conditioning captured CS–US associations without a response contingency (e.g., aversive conditioning). Third, declarative and explicit memory encompassed knowledge- or memory-based tasks without trial-by-trial feedback (e.g., item recognition, explicit rule knowledge). (ii) Type of stressor: Guided by theory on social-evaluative threat the stress paradigms were aggregated into two categories: psychosocial stressors (TSST, SECPT, MAST) and primarily physiological stressors (CPT, auditory stimulation). This approach addressed the unstable estimates and low power due to very few studies per category. (iii) Stressor time: The time at which the Stressor was presented to the participants was coded as the three categories prelearning, concurrent and postlearning. Lastly, because there was no variability in the primary analysis the variable (iv) study design was only included in the secondary analysis: The variable was analysed as a possible moderator due to the possibility of more similar data in within-subject compared to between-subject designs. To analyse a potential moderating effect of these variables, an omnibus test was performed. All categorical moderators were included in the model as dummy coded variables.

Sensitivity Analyses

To assess the robustness of the findings, a sensitivity analysis was conducted in which effect sizes from studies, that reported non-significant results without numeric values or only p-values, were conservatively set to zero. The meta-analysis was re-estimated under this worst-case assumption to assess the stability of the main effects.

Publication Bias

There is evidence that studies that report significant results are more likely to be published than studies that report non-significant findings (Borenstein, 2021). To determine whether confounding effects of publication bias were present, funnel plots were used to visually assess potential asymmetry in study publication.

Furthermore, tests to assess a possible publication bias that are suitable for dependent effect sizes, respectively multi-level models, were used. PET-PEESE meta-regression (Stanley & Doucouliagos, 2014) analyses were conducted to estimate a bias-corrected version of the pooled effect. The precision-effect test (PET) includes the standard error as a predictor and therefor controls for small study effects. Compared to the precision-effect estimate with standard error (PEESE) it is a more conservative approach that tests whether true effects equal zero. The PEESE methods uses the effect size variance as a predictor and estimates the true effects adjusted for small-study bias. PET works best when the true effect is zero, and PEESE performs better when the true effect is not zero. To balance out the individual strengths of both tests, it is recommended to combine the methods (Harrer et al., 2021).

To assess the robustness of the overall effect a Robust Variance Estimation (RVE) model was fitted. Although RVE is not a formal publication-bias test, it serves as a robustness test. If the overall effect remains stable under RVE, the findings are unlikely to be driven by small-study effects or model misspecification (Fisher & Tipton, 2015).

Results

Study Inclusion

The process of study selection is illustrated in a PRISMA diagram (Figure 1). The initial database search yielded 348 records after removal of duplicates. First, all abstracts were screened, and relevant articles were selected for full-text assessment. This resulted in 70 potentially relevant articles, which were assessed for eligibility against the inclusion criteria. Ultimately, 30 studies met the criteria and were included in the meta-analysis. The included studies were published between the years 2010 and 2025 in the following countries: Germany (n = 9), USA (n = 7), Netherlands (n = 6), Australia (n = 2), Portugal (n = 2), Argentina (n = 1), Canada (n = 1), Spain (n = 1), United Kingdom (n = 1).

Main Results

The random-effects multilevel model (REML) of the primary analysis reveals a small negative overall effect, g = -0.31, 95% CI [-0.48, -0.14], p < .001. The heterogeneity test

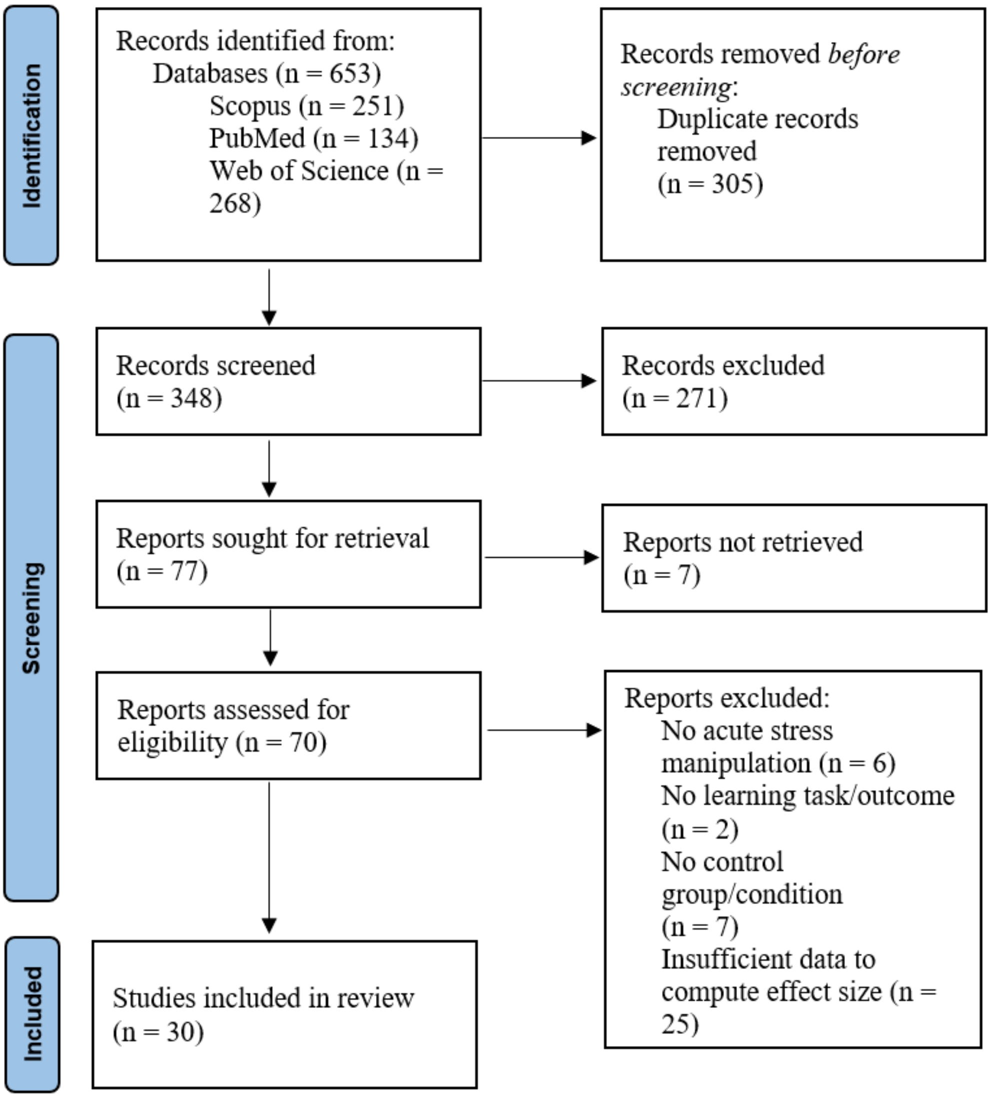

Figure 1: PRISMA flowchart for selection process

indicates significant variability among effect sizes, $Q(28) = 48.81$, $p = .009$. The dataset consists of $k = 29$ effect sizes extracted from 17 studies.

In comparison the REML of the secondary analysis can be interpreted as a large negative overall effect, $r = -0.51$, 95% CI [-0.58, -0.43], $p < .0001$. In alignment with the primary analysis the heterogeneity test of the secondary analysis shows significant variability among effect sizes as well, $Q(24) = 55.69$, $p = < .001$. The dataset includes $k = 25$ effect sizes from 19 studies.

The explorative analysis of goal directed versus habitual behaviour with a dataset of $k = 7$ effect sizes from 6 studies shows a medium negative overall effect, $r = -0.43$, 95% CI [-0.62, -0.20], $p < .0001$. The heterogeneity test was significant, $Q(6) = 24.13$, $p = < .001$.

Table 1. Three-level meta-analysis: Main effects

	k (ES)	*estimate*	*p*-values	95% CI		*Q(df)*
Primary analysis	17 (29)	-0.31(*g*)	<.001	-0.48	-0.14	48.81(29)
Secondary analysis	19 (25)	-0.51(*r*)	<.0001	-0.58	-0.43	55.69(24)
Goal vs. habit	6 (7)	-0.43(*r*)	<.0001	-0.62	-0.20	24.13(6)

Note: k = Studies; ES = Effect sizes; *g* = Hedges‘ *g*; *r = Pearson correlation* 95% CI = Confidence interval; Q(df) = Heterogeneity

Heterogeneity Analyses

To further investigate the variability among effect sizes indicated by the significant residual heterogeneity test, models with and without each variance component as well as the distribution over the three levels was analysed.

Primary analysis

Removing the within-study variance (level-2) did not worsen the model fit (p = 1.00; AIC reduced = 22.00 vs. full = 24.00), indicating a negligible level-2 component. In contrast, removing the between-study variance (level-3) reduced model fit (p = .076; AIC reduced = 25.15 vs. full = 24.00), consistent with meaningful heterogeneity across studies. The total variance is distributed across all three levels of the meta-analytic model as follows: 52.67 percent of the total variance can be attributed to level-1 variance, 4.06 percent to differences between effect sizes within studies at level 2, and 47.33 percent to differences between studies at level-3.

Secondary analysis

The within-study component was again unsupported (p = 1.00; AIC reduced = -5.62 vs. full = -3.62). Removing the between-study component significantly worsened the model fit (p = .008; AIC reduced = 1.42 vs. full = -3.62), indicating clear between-study heterogeneity. The variance split was 37.86 percent sampling error, 4.83 percent within-study, and 62.14 percent between-study variance.

Moderator Analyses

Learning type as moderator

Compared to instrumental and RL tasks (reference) neither declarative and explicit memory (primary analysis: Δg = 0.14, p = .582; secondary analysis: Δr = -0.21, p = .177) nor Pavlovian tasks (Δg = -0.14, p = .619; secondary analysis: Δr = -0.08, p = .733) differed significantly in either analysis.

Stress time as moderator

Postlearning stress had no significant moderating effect when compared to prelearning stress (reference) in either analysis (primary analysis: Δg = -.00, p = .996; secondary analysis: Δr = -0.01, p = .928). A significant effect of concurrent stress was only found in the secondary analysis (primary analysis: Δg = -0.02, p = .953; secondary analysis: Δr = 0.32, p = .015). Which would translate to a more severe impairing effect of pre-learning compared to concurrent stress.

Stressor type as moderator

There was no effect found when comparing stressors with a psychosocial component (reference) to physiological stressors (primary analysis: Δg = -.12, p = .481; secondary analysis: Δr = 0.11, p = .332).

Study design as a moderator

Secondary analysis showed a significant effect of within-subject (reference) compared to between-subject designs (Δr = -0.20, p = .037) indicating that the impairing effect of stress was significantly stronger in studies which used within-subject designs.

Goal-directed versus habitual behaviour

In the subset of studies directly testing goal-directed versus habitual control the multilevel random-effects meta-analysis (REML) shows a negative overall effect, r = -0.43, 95% CI [-0.62, -0.20], $p < .0001$. The heterogeneity test indicates significant variability among effect sizes, $Q(6)$ = 24.13, p = < .001. The dataset consists of k = 7 effect sizes extracted from 6 studies.

Sensitivity Analyses

When non-significant effects were conservatively set to zero and included in the primary model, the overall effect was still small and negative, g = -0.29, 95% CI [-0.46, -0.13],

$p < .001$. The heterogeneity test still indicates significant variability among effect sizes, $Q(30) = 54.01$, $p = .005$. The dataset now consists of $k = 31$ effect sizes from 17 studies.

Including non-significant results in the secondary analysis reduced the overall effect to a medium effect, $r = -0.44$, 95% CI [-0.54, -0.34], $p < .0001$. Notably, the variability among effect sizes increased with the inclusion, $Q(29) = 113.43$, $p < .0001$. This analysis was based on $k = 30$ effect sizes from 23 studies.

Across both the primary and secondary models, the sensitivity analyses produced effect estimates that remained negative and statistically significant, indicating that the overall pattern of results is not dependent on selective reporting of significant effects. Although the effect sizes were somewhat attenuated when conservatively treating non-significant or unreported results as zeros, the core conclusions did not change. This suggests a reasonable degree of robustness, while also highlighting that the magnitude of the estimated effects is sensitive to assumptions about missing information.

Publication Bias

The visual assessment of the funnel plots of the individual effect sizes and their standard errors (figure 3) provides a first impression of a possible small-study effects. *Primary analysis*: Most studies cluster in the middle range of standard errors, which indicates that the studies have similar levels of precision. The overall distribution of points appears roughly symmetric around the pooled effect estimate, although some scatter is visible on both sides. Importantly, there is no clear pattern of asymmetry. *Secondary analysis*: The points are scattered around the pooled effect estimate, with visible dispersion on both the left and right sides. Both negative and positive effect sizes are represented across the range of standard errors, and there is no obvious absence of studies on one specific side of the funnel. The slight asymmetry present is modest and can be attributed to sampling variability, heterogeneity across studies, and the dependence structure of the data. Overall, the plots do not provide compelling visual evidence for strong small-study effects or publication bias. Given the moderate number of studies and the presence of multilevel dependence in the data, the funnel plot should be interpreted only as a descriptive visual tool.

Potential small-study effects were examined using PET–PEESE meta-regressions based on Hedges' g from the primary analysis and Fisher's z for the secondary analysis. *Primary model:* In the PET model the slope of the standard error was not significant, $b = 2.23$, $SE = 1.83$, $t(27) = 1.49$, $p = .233$, indicating no evidence of a linear association

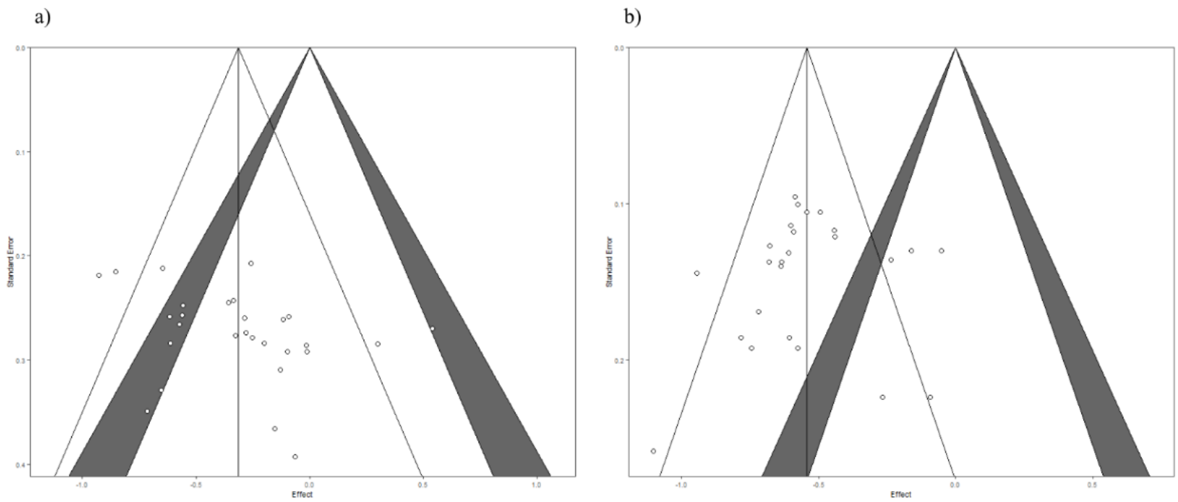

Figure 2: Funnel plot of effect sizes by standard error for a) primary analysis and b) secondary analysis

between study precision and effect magnitude. The PET intercept was negative but not statistically significant, b = -0.92, SE = 0.50, t(27) = -1.83, p = .079. The PEESE model, yielded a non-significant moderator term, b = 3.21, SE = 3.18, t(27) = 1.01, p = .322. However, the PEESE intercept was significantly negative, b = -0.56, SE = 0.26, t(27) = -2.18, p = .038, indicating that the bias-adjusted underlying effect remained negative. *Secondary model*: In the PET model the slope was not significant as well, b = -1.24, SE = 1.37, t(23) = -0.90, p = .377. The PET intercept was negative but not statistically significant, b = -0.38, SE = 0.20, t(23) = -1.91, p = .069. When transformed to the r metric, this corresponded to a small negative effect (r = -0.37). In the PEESE model, the moderator term was not significant, b = -3.81, SE = 4.91, t(23) = -0.93, p = .361. The PEESE intercept was significantly negative, b = -0.48, SE = 0.10, t(23) = -4.71, $p < .001$, indicating a robust negative underlying effect after accounting for potential small-study bias. The corresponding estimate in the r metric was r = -0.45. Taken together, PET–PEESE analyses did not provide evidence of small-study bias,

To further assess possible biases the robust variance estimation model was conducted. The model indicates a significant negative overall effect, r = -0.50, 95% CI [-0.58, -0.40], $p < .001$. This suggests that the overall association remained substantial and statistically significant when accounting for dependence among effect sizes.

Discussion

The present meta-analysis examined the effect of acute stress on learning using a three-level model that accounted for dependence among effect sizes both within and across studies. Analyses were conducted separately for studies reporting simple group contrasts (stress vs. control) and studies reporting interaction effects. Across both analytical approaches, the results indicated a small to medium impairing effect of acute stress on learning performance. This aligns with theoretical assumptions that stress can interfere with learning-related cognitive processes, although the magnitude of impairment varied across study designs and stress characteristics.

Only limited evidence emerged for moderating influences. In the secondary analysis, concurrent stress was associated with a less pronounced negative effect compared to pre-learning stress, suggesting that the timing of stress induction may meaningfully shape learning outcomes. Within-subject designs also showed a significantly stronger detrimental stress effect compared to between-subject designs, potentially reflecting increased sensitivity to stress-induced performance fluctuations when individual differences are controlled. It must be noted that learning types as well as stressor types were grouped based on theoretical considerations, but the available number of studies did not permit fine-grained differentiation. A larger body of evidence may reveal more specific moderating influences of learning mechanisms and stressor types.

Publication-bias diagnostics consistently indicated that the observed effects were unlikely to be driven by small-study effects or selective reporting. PET–PEESE analyses showed no significant association between effect size and study precision, and the bias-adjusted PEESE intercept remained significantly negative. Because PET–PEESE can be sensitive to high heterogeneity and relatively small numbers of studies, robust variance estimation (RVE) was additionally employed (Harrer et al., 2021). The RVE estimate closely matched the three-level model and remained significantly negative, reinforcing the conclusion that the overall effect is robust to violations of independence assumptions and not substantially influenced by publication bias. Although no indications of bias were detected analytically or visually, it was noticeable that many studies did not report full statistical estimates for non-significant results, relying instead on threshold-based p-value statements. This pattern underscores the importance of transparent reporting practices in future research.

Overall, the evidence supports the conclusion that acute stress negatively affects learning mechanisms. However, the effect is not universally “impairing” across all forms of learning. Exploratory comparisons between goal-directed and habitual behavior highlight that stress-induced shifts toward habitual responding can be functional and adaptive in certain contexts. This suggests that the relationship between acute stress and learning outcomes is more nuanced than a simple beneficial versus detrimental dichotomy. The stronger effects observed in the secondary model that incorporated interaction terms also point toward a more complex underlying structure. Future research should further investigate these dynamics, including potential cumulative effects of repeated stress exposure and the distinction between acute and chronic stress. More comprehensive moderator analyses, enabled by a larger pool of studies and improved reporting standards, will be essential to fully understand how stress interacts with different learning systems.

Limitations

Several limitations warrant consideration. First, the total number of available studies, although adequate for meta-analytic modelling, remains relatively small for more sophisticated moderator analyses, particularly when distinguishing among different learning mechanisms and stressor types. Some theoretically meaningful categories may have lacked sufficient statistical power to detect moderation effects. Second, the primary literature showed inconsistencies in reporting practices; many studies did not provide full statistical estimates for non-significant results, limiting precision in effect-size extraction and potentially masking subtle patterns. Third, substantial methodological heterogeneity across studies - such as differences in stress induction methods, learning tasks, sample characteristics, and timing of stress exposure - may have introduced variability that could not be fully accounted for by the available moderators. Although the three-level model addresses dependencies among effect sizes, residual heterogeneity remained significant, indicating that unmeasured study-level factors likely contributed to variability in outcomes. Finally, the meta-analysis focused exclusively on acute stress. The potential cumulative or interactive effects of repeated or chronic stress could not be evaluated with the existing dataset.

References

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Bali, A., & Jaggi, A. S. (2015). Clinical experimental stress studies: Methods and assessment. *Reviews in the Neurosciences*, *26*(5), 555–579. https://doi.org/10.1515/revneuro-2015-0004

Berghorst, L. H., Bogdan, R., Frank, M. J., & Pizzagalli, D. A. (2013). Acute stress selectively reduces reward sensitivity. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00133

Bogdan, R., Santesso, D. L., Fagerness, J., Perlis, R. H., & Pizzagalli, D. A. (2011). Corticotropin-releasing hormone receptor type 1 (CRHR1) genetic variation and stress interact to influence reward learning. *Journal of Neuroscience*, *31*(37), 13246–13254. Scopus. https://doi.org/10.1523/JNEUROSCI.2661-11.2011

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (Second edition). Wiley.

Carvalheiro, J., Conceição, V. A., Mesquita, A., & Seara-Cardoso, A. (2021). Acute stress impairs reward learning in men. *Brain and Cognition*, *147*. Scopus. https://doi.org/10.1016/j.bandc.2020.105657

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed., reprint). Psychology Press.

Fisher, Z., & Tipton, E. (2015). *robumeta: An R-package for robust variance estimation in meta-analysis* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1503.02220

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide* (1. Aufl.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

Nitschke, J. P., Giorgio, L.-M., Zaborowska, O., & Sheldon, S. (2020). Acute psychosocial stress during retrieval impairs pattern separation processes on an episodic memory task. *Stress*, *23*(4), 437–443. https://doi.org/10.1080/10253890.2020.1724946

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Schwabe, L. (2017). Memory under stress: From single systems to network changes. *European Journal of Neuroscience*, *45*(4), 478–489. https://doi.org/10.1111/ejn.13478

Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1740–1749. https://doi.org/10.1016/j.neubiorev.2011.07.002

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, *29*(22), 7191–7198. Scopus. https://doi.org/10.1523/JNEUROSCI.0979-09.2009

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias: T. D. STANLEY AND H. DOUCOULIAGOS. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience*, *10*(6), 397–409. https://doi.org/10.1038/nrn2647

Viechtbauer, W. (2010). Conducting Meta-Analyses in *R* with the **metafor** Package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Appendix

Zusammenfassung

Akuter Stress beeinflusst grundlegende kognitive Prozesse, doch seine Auswirkungen auf Lernen werden in empirischen Studien uneinheitlich berichtet. Diese Arbeit untersucht den Zusammenhang systematisch und quantitativ mittels einer Drei-Ebenen-Metaanalyse experimenteller akuter Stressparadigmen. Eine umfassende Literatursuche identifizierte 30 Studien mit erwachsenen Teilnehmenden, die akuten Stressoren ausgesetzt waren und lernbezogene Outcomes berichteten. Die daraus gewonnenen Effektgrößen wurden mithilfe von Random-Effects-Mehrebenenmodellen analysiert, die Abhängigkeiten innerhalb von Studien berücksichtigen.

Insgesamt zeigt sich ein kleiner bis mittlerer beeinträchtigender Effekt von akutem Stress auf die Lernleistung. Likelihood-Ratio-Tests ergaben keine Evidenz für Heterogenität innerhalb der Studien, während signifikante Heterogenität zwischen den Studien gefunden wurde. Moderatoranalysen zeigten keine Unterschiede bei Lernarten, Stressortypen oder Stress nach dem Lernen. Im Gegensatz dazu war zeitgleicher Stress weniger beeinträchtigend als Stress vor dem Lernen, und Within-Subject-Designs zeigten stärkere negative Effekte als Between-Subject-Designs.

Analysen zum Publikationsbias lieferten keinen Hinweis auf Small-Study-Effekte und bestärkten die Annahme eines zugrunde liegenden beeinträchtigenden Effekts. Die Ergebnisse zeigen, dass akuter Stress Lernen beeinträchtigen kann, insbesondere wenn er der Lernphase vorausgeht und intraindividuelle Vergleiche möglich sind, und betonen die Bedeutung methodischer und kontextueller Faktoren.

Study characteristics: a) Primary analysis; b) secondary

a)

studyID	effectsizeID	Study_plot	Year_of_Publishing	publicationstatus	Peer_Review	Country	Group_Allocation	Age_N1_Stress	Age_N2_Control	male	female	Gender	N1	NC	N
01	01	Beck et al., 2022	2022	yes	yes	Germany	R	24.85	25.81	52	0	m	26	26	52
01	02	Beck et al., 2022	2022	yes	yes	Germany	R	24.85	25.81	52	0	m	26	26	52
01	03	Beck et al., 2022	2022	yes	yes	Germany	R	24.85	25.81	52	0	m	26	26	52
02	04	Boyle et al., 2020	2020	yes	yes	USA	R	19.81	20.71	0	44	f	31	13	44
02	05	Boyle et al., 2020	2020	yes	yes	USA	R	19.81	20.72	0	51	f	34	17	51
03	06	Carvalheiro et al., 2021	2021	yes	yes	Portugal	NA	21.90	21.90	62	0	m	NA	NA	62
03	07	Carvalheiro et al., 2021	2021	yes	yes	Portugal	NA	21.90	21.90	62	0	m	NA	NA	62
04	08	de Berker et al., 2016	2016	yes	yes	United Kingdom	R	NA	NA	NA	NA	m/f	26	30	56
04	09	de Berker et al., 2016	2016	yes	yes	United Kingdom	R	NA	NA	NA	NA	m/f	25	30	55
05	10	Dunsmoor et al., 2017	2017	yes	yes	USA	NA	23.00	20.00	NA	NA	m/f	20	21	41
06	11	Goldfarb et al., 2017	2017	yes	yes	USA	R	21.97	22.53	26	34	m/f	27	27	54
06	12	Goldfarb et al., 2017	2017	yes	yes	USA	R	21.97	22.53	26	34	m/f	27	27	54
07	13	Hartogsveld et al., 2020	2020	yes	yes	Netherlands	semi-random	NA	NA	18	47	m/f	30	35	65
07	14	Hartogsveld et al., 2020	2020	yes	yes	Netherlands	semi-random	NA	NA	18	47	m/f	30	35	65
08	15	Kamp et al., 2019	2019	yes	yes	Germany	NA	23.35	23.35	37	0	m	19	18	37
09	16	Lemmens et al., 2022	2022	yes	yes	Netherlands	R	NA	NA	19	42	m/f	31	30	61
09	17	Lemmens et al., 2022	2022	yes	yes	Netherlands	R	NA	NA	19	42	m/f	31	30	61
10	18	Lighthall et al., 2013	2013	yes	yes	USA	R	47.80	47.95	48	48	m/f	48	48	96
10	19	Lighthall et al., 2013	2013	yes	yes	USA	R	47.80	47.95	48	48	m/f	48	48	96
10	20	Lighthall et al., 2013	2013	yes	yes	USA	R	47.80	47.95	48	48	m/f	48	48	96
11	21	Paul et al., 2018	2018	yes	yes	Germany	R	24.80	25.60	32	0	f	16	16	32
12	22	Pritchard et al., 2018	2018	yes	yes	Australia	R	22.00	23.00	8	62	m/f	35	35	70
12	23	Pritchard et al., 2018	2018	yes	yes	Australia	R	22.00	23.00	8	62	m/f	35	35	70
13	24	Quaedflieg et al., 2019	2019	yes	yes	Netherlands/Germany/UK	R			56	56	m/f	47	48	95
14	25	Raio et al., 2020	2020	yes	yes	USA	R	25.67	25.67	26	30	m/f	20	28	49
14	26	Raio et al., 2020	2020	yes	yes	USA	R	25.67	25.67	26	30	m/f	20	28	49
15	27	Smeets et al., 2019	2019	yes	yes	Netherlands	R	NA	NA	20	52	m/f	35	36	71
16	28	Vogel & Schwabe, 2018	2018	yes	yes	Germany	NR	NA	NA	29	30	m/f	30	29	59
17	29	Wieland et al., 2023	2023	yes	yes	Germany	NA	NA	NA	28	0	m	NA	NA	28

b)

studyID	effectsizeID	Study_plot	Year_of_Publishing	publicationstatus	Peer_Review	Country	Group_Allocation	Age_N1_Stress	Age_N2_Control	male	female	Gender	N1	NC	N
01	01	Bogdan et al., 2011	2011	yes	yes	USA	NA	21.88	21.88	0.00	75.00	f	NA	NA	75.00
02	02	Carvalheiro et al., 2021	2021	yes	yes	Portugal	NA	21.90	21.90	62.00	0.00	m	NA	NA	62.00
02	03	Carvalheiro et al., 2021	2021	yes	yes	Portugal	NA	21.90	21.90	62.00	0.00	m	NA	NA	62.00
03	04	Carvalheiro et al., 2022	2022	yes	yes	Portugal	NA	23.00	23.00	23.00	0.00	m	NA	NA	23.00
03	05	Carvalheiro et al., 2022	2022	yes	yes	Portugal	NA	23.00	23.00	23.00	0.00	m	NA	NA	23.00
04	06	Ehlers & Todd, 2017	2017	yes	yes	Canada	R	21.30	21.00	NA	NA	m/f	NA	NA	102.00
04	07	Ehlers & Todd, 2017	2017	yes	yes	Canada	R	21.30	21.00	NA	NA	m/f	NA	NA	112.00
05	08	Goldfarb et al., 2017	2017	yes	yes	USA	R	21.97	22.53	26.00	34.00	m/f	27.00	27.00	54.00
06	09	Goldfarb et al., 2018	2018	yes	yes	USA	within-subjects	23.40	23.40	13.00	17.00	m/f	NA	NA	30.00
06	10	Goldfarb et al., 2018	2018	yes	yes	USA	within-subjects	23.40	23.40	13.00	17.00	m/f	NA	NA	30.00
07	11	Hartogsveld et al., 2020	2020	yes	yes	Netherlands	semi-random assignment	NA	NA	18.00	47.00	m/f	30.00	35.00	65.00
08	12	Kruse et al., 2018	2018	yes	yes	Germany	NA	23.48	23.83	56.00	0.00	m	27.00	29.00	56.00
08	13	Kruse et al., 2018	2018	yes	yes	Germany	NA	23.48	23.83	56.00	0.00	m	27.00	29.00	56.00
09	14	Larrosa et al., 2017	2017	yes	yes	Argentina	R	23.40	23.40	6.00	12.00	m/f	8.00	10.00	18.00
10	15	Lemmens et al., 2022	2022	yes	yes	Netherlands	R	NA	NA	19.00	42.00	m/f	31.00	30.00	61.00
11	16	Molins et al., 2025	2025	yes	yes	Spain	R	21.85	20.58	15.00	58.00	m/f	37.00	39.00	76.00
12	17	Paul et al., 2018	2018	yes	yes	Germany	R	24.80	25.60	32.00	0.00	f	16.00	16.00	32.00
13	18	Paul et al., 2020	2020	yes	yes	Germany	R	NA	NA	38.00	0.00	m	17.00	21.00	38.00
14	19	Radenbach et al., 2015	2015	yes	yes	Germany	NA	NA	NA	39.00	0.00	m	NA	NA	32.00
15	20	Schwabe & Wolf, 2010	2010	yes	yes	Germany	R	23.40	23.40	28.00	23.00	m/f	28.00	23.00	51.00
16	21	Smeets et al., 2018	2018	yes	yes	Netherlands	R	NA	NA	NA	NA	m/f	NA	NA	71.00
17	22	van Ruitenbeek et al., 2021	2021	yes	yes	Netherlands	R	22.51	22.60	36.00	57.00	m/f	45.00	48.00	93.00
17	23	van Ruitenbeek et al., 2021	2021	yes	yes	Netherlands	R	22.51	22.60	36.00	57.00	m/f	45.00	48.00	93.00
18	24	Voulgaropoulou et al., 2022	2022	yes	yes	Netherlands	R	23.48	23.80	37.00	43.00	m/f	40.00	40.00	80.00
19	25	Walter et al., 2021	2021	yes	yes	Australia	NR	22.43	22.43	19.00	38.00	m/f	NA	NA	57.00

Moderators: Primary analysis; b) secondary analysis

a)

studyID	effectsizeID	Study_plot	Comparison_group	Stressor	Stressor_psychosocial	Stressor_physiological	Learning_Type	Instrumental_RL	Declarative_Explicit	Pavlovian	Task	stress_time_	pre	post	con.
01	01	Beck et al., 2022	between-subjects	SECPT	1	0	Instrumental/RL	1	0	0	instrumental counterconditioning task	postlearning	0	1	0
01	02	Beck et al., 2022	between-subjects	SECPT	1	0	Instrumental/RL	1	0	0	instrumental counterconditioning task	postlearning	0	1	0
01	03	Beck et al., 2022	between-subjects	SECPT	1	0	Instrumental/RL	1	0	0	instrumental counterconditioning task	postlearning	0	1	0
02	04	Boyle et al., 2020	between-subjects	TSST	1	0	Instrumental/RL	1	0	0	PRT	postlearning	0	1	0
02	05	Boyle et al., 2020	between-subjects	TSST	1	0	Instrumental/RL	1	0	0	Effort expenditure for rewards task (EEfRT)	postlearning	0	1	0
03	06	Carvalheiro et al., 2021	between-subjects	auditory stimulus	0	1	Instrumental/RL	1	0	0	Reinforcement-learning task	concurrent	0	0	1
03	07	Carvalheiro et al., 2021	between-subjects	auditory stimulus	0	1	Instrumental/RL	1	0	0	Reinforcement-learning task	concurrent	0	0	1
04	08	de Berker et al., 2016	between-subjects	SECPT	1	0	Instrumental/RL	1	0	0	Go/NoGo	prelearning	1	0	0
04	09	de Berker et al., 2016	between-subjects	SECPT	1	0	Instrumental/RL	1	0	0	Go/NoGo	prelearning	1	0	0
05	10	Dunsmoor et al., 2017	between-subjects	CPT	0	1	Pavlovian	0	0	1	Delayed Test of Generalization	prelearning	1	0	0
06	11	Goldfarb et al., 2017	between-subjects	CPT	0	1	Declarative/Explicit	0	1	0	Search task	postlearning	0	1	0
06	12	Goldfarb et al., 2017	between-subjects	CPT	0	1	Declarative/Explicit	0	1	0	Search task	postlearning	0	1	0
07	13	Hartogsveld et al., 2020	between-subjects	MAST	1	0	Instrumental/RL	1	0	0	Slips-of-Action	prelearning	1	0	0
07	14	Hartogsveld et al., 2020	between-subjects	MAST	1	0	Instrumental/RL	1	0	0	Slips-of-Action	prelearning	1	0	0
08	15	Kamp et al., 2019	between-subjects	TSST	1	0	Declarative/Explicit	0	1	0	Item recognition	prelearning	1	0	0
09	16	Lemmens et al., 2022	between-subjects	MAST	1	0	Pavlovian	0	0	1	Avoidance - reversal-learning	prelearning	1	0	0
09	17	Lemmens et al., 2022	between-subjects	MAST	1	0	Pavlovian	0	0	1	Avoidance - reversal-learning	prelearning	1	0	0
10	18	Lighthall et al., 2013	between-subjects	CPT	0	1	Instrumental/RL	1	0	0	Probabilistic selection task	prelearning	1	0	0
10	19	Lighthall et al., 2013	between-subjects	CPT	0	1	Instrumental/RL	1	0	0	Probabilistic selection task	prelearning	1	0	0
10	20	Lighthall et al., 2013	between-subjects	CPT	0	1	Instrumental/RL	1	0	0	Probabilistic selection task	prelearning	1	0	0
11	21	Paul et al., 2018	between-subjects	SECPT	1	0	Instrumental/RL	1	0	0	Category learning task	prelearning	1	0	0
12	22	Pritchard et al., 2018	between-subjects	Negative emotional appraisal	0	1	Instrumental/RL	1	0	0	Instrumental outcome devaluation	concurrent	0	0	1
12	23	Pritchard et al., 2018	between-subjects	Negative emotional appraisal	0	1	Instrumental/RL	1	0	0	Instrumental outcome devaluation	concurrent	0	0	1
13	24	Quaedflieg et al., 2019	between-subjects	MAST	1	0	Instrumental/RL	1	0	0	Instrumental outcome devaluation	postlearning	0	1	0
14	25	Raio et al., 2020	between-subjects	CPT	0	1	Instrumental/RL	1	0	0	sequential decision-making	prelearning	1	0	0
14	26	Raio et al., 2020	between-subjects	CPT	0	1	Instrumental/RL	1	0	0	sequential decision-making	prelearning	1	0	0
15	27	Smeets et al., 2019	between-subjects	MAST	1	0	Instrumental/RL	1	0	0	slips-of-action	postlearning	0	1	0
16	28	Vogel & Schwabe, 2018	between-subjects	SECPT	1	0	Declarative/Explicit	0	1	0	explicit rule knowledge test	prelearning	1	0	0
17	29	Wieland et al., 2023	between-subjects	TSST	1	0	Instrumental/RL	1	0	0	Probabilistic reversal learning	prelearning	1	0	0

b)

studyID	effectsizeID	Study_plot	Comparison_group	within_design	between_design	Stressor	Stressor_psychosocial	Stressor_physiological	Learning_Type	Instrumental_RL	Declarative_Explicit	Pavlovian	Task	stress time	pre	post	concurrent
01	01	Bogdan et al., 2011	within-subjects	1	0	Threat of shock	0	1	RL	1	0	0	Probabilistic Reward Task	concurrent	0	0	1
02	02	Carvalheiro et al., 2021	within-subjects	1	0	auditory stimulus	0	1	Instrumental	1	0	0	Reinforcement-learning task	concurrent	0	0	1
02	03	Carvalheiro et al., 2021	within-subjects	1	0	auditory stimulus	0	1	Instrumental	1	0	0	Reinforcement-learning task	concurrent	0	0	1
03	04	Carvalheiro et al., 2022	within-subjects	1	0	auditory stimulus	0	1	RL	1	0	0	Reinforcement-learning task	concurrent	0	0	1
03	05	Carvalheiro et al., 2022	within-subjects	1	0	auditory stimulus	0	1	RL	1	0	0	Reinforcement-learning task	concurrent	0	0	1
04	06	Ehlers & Todd, 2017	between-subjects	0	1	SECPT	1	0	Instrumental	1	0	0	Operant task	prelearning	1	0	0
04	07	Ehlers & Todd, 2017	between-subjects	0	1	SECPT	1	0	Instrumental	1	0	0	Operant task	prelearning	1	0	0
05	08	Goldfarb et al., 2017	between-subjects	0	1	CPT	0	1	Memory	0	1	0	Search task	postlearning	0	1	0
06	09	Goldfarb et al., 2018	within-subjects	1	0	CPT	0	1	Memory	0	1	0	Item/Associative recognition	prelearning	1	0	0
06	10	Goldfarb et al., 2018	within-subjects	1	0	CPT	0	1	Memory	0	1	0	Item/Associative recognition	postlearning	0	1	0
07	11	Hartogsveld et al., 2020	between-subjects	0	1	MAST	1	0	Instrumental	1	0	0	Slips-of-Action	prelearning	1	0	0
08	12	Kruse et al., 2018	between-subjects	0	1	TSST	1	0	RL	1	0	0	Modified Monetary Incentive Delay	postlearning	0	1	0
08	13	Kruse et al., 2018	between-subjects	0	1	TSST	1	0	RL	1	0	0	Modified Monetary Incentive Delay	postlearning	0	1	0
09	14	Larrosa et al., 2017	between-subjects	0	1	CPT	0	1	Memory	0	1	0	Cued recall of nonsense syllables	postlearning	0	1	0
10	15	Lemmens et al., 2022	between-subjects	0	1	MAST	1	0	Pavlovian	0	0	1	Avoidance - reversal-learning	postlearning	0	1	0
11	16	Molins et al., 2025	between-subjects	0	1	TSST	1	0	Decision	1	0	0	Iowa Gambling Task	postlearning	0	1	0
12	17	Paul et al., 2018	within-subjects	1	0	SECPT	1	0	Category	1	0	0	Category learning task	prelearning	1	0	0
13	18	Paul et al., 2020	between-subjects	0	1	SECPT	1	0	RL	1	0	0	Probabilistic reward learning	prelearning	1	0	0
14	19	Radenbach et al., 2015	within-subjects	1	0	TSST	1	0	RL	1	0	0	Two-step reinforcement learning task	prelearning	1	0	0
15	20	Schwabe & Wolf, 2010	between-subjects	0	1	SECPT	1	0	Instrumental	1	0	0	extinction test	prelearning	1	0	0
16	21	Smeets et al., 2018	within-subjects	1	0	MAST	1	0	Instrumental	1	0	0	slips-of-action	postlearning	0	1	0
17	22	van Ruitenbeek et al., 2021	between-subjects	0	1	MAST	1	0	Instrumental	1	0	0	Devaluation	prelearning	1	0	0
17	23	van Ruitenbeek et al., 2021	between-subjects	0	1	MAST	1	0	Instrumental	1	0	0	Devaluation	prelearning	1	0	0
18	24	Voulgaropoulou et al., 2022	within-subjects	1	0	MAST	1	0	RL	1	0	0	cost-benefit reinforcement learning paradigm	prelearning	1	0	0
19	25	Walter et al., 2021	within-subjects	1	0	TSST	1	0	RL	1	0	0	Spatialtrial-and-errorlearning	prelearning	1	0	0

Comparison and statistics: a) Primary analysis; b) secondary

a)

study	effect	Authors	Study_plot	Comparison_specific	statistics_for_comparison	M1_pr	MC_pr	M1_po	MC_po	SD1_	SD2_	SD_p	F	df1	df2	d	g	var_g	r	t	beta	SE	Chi-square
01	01	Beck et al.	Beck et al., 2022	stress vs. control - positive consequences	press the risk key	-	-	-	-	-	-	-	-	-	-	0.30	0.30	0.08	0.15	-	0.55	-	-
01	02	Beck et al.	Beck et al., 2022	stress vs. control - negative consequences	press the risk key	-	-	-	-	-	-	-	-	-	-	0.30	0.30	0.08	0.15	-	0.55	-	-
01	03	Beck et al.	Beck et al., 2022	stress vs. Control - counterconditioning positive	press the risk key	-	-	-	-	-	-	-	-	-	-	-0.20	-0.20	0.08	-0.10	-	-0.37	-	-
02	04	Boyle et al	Boyle et al., 2020	Stress vs. Control Response bias	PRT response bias score	-	-	-	-	-	-	-	-	-	-	0.13	-0.13	0.10	-	-	0.24	0.22	-
02	05	Boyle et al	Boyle et al., 2020	Stress vs. Control EEfRT reward motivation	EEfRT reward motivation	-	-	-	-	-	-	-	-	-	-	0.02	-0.02	0.08	-	-	0.03	0.04	-
03	06	Carvalheiro et al.	Carvalheiro et al., 2021	Positive learning rate (α+)	Accuracy	-	-	-	-	-	-	-	-	-	61	-0.29	-0.29	0.07	-	-	-	-	-
03	07	Carvalheiro et al.	Carvalheiro et al., 2021	Negative learning rate (α-)	Accuracy	-	-	-	-	-	-	-	-	-	61	-0.09	-0.09	0.07	-	-	-	-	-
04	08	de Berker et al.	de Berker et al., 2016	Go to Win (Stress vs. Control)	% of go choices	-	-	-	-	-	-	-	-	-	-	0.28	-0.28	0.07	-	-	-	-	4.20
04	09	de Berker et al.	de Berker et al., 2016	Go to Avoid Losing (Stress vs. Control)	% of go choices	-	-	-	-	-	-	-	-	-	-	0.33	-0.33	0.08	-	-	-	-	5.45
05	10	Dunsmoor et al.	Dunsmoor et al., 2017	main effect group	shock expectancy	-	-	-	-	-	-	-	4.55	1	39	0.67	-0.65	0.11	-	2.13	-	-	-
06	11	Goldfarb et al.	Goldfarb et al., 2017	stress vs. Control	Accuracy	96.98	97.58	98.40	97.93	1.70	1.74	1.72	-	-	-	0.62	-0.61	0.08	-	-	-	-	-
06	12	Goldfarb et al.	Goldfarb et al., 2017	stress vs. Control	RT	1.42	1.33	1.14	1.11	0.22	0.25	0.24	-	-	-	-0.25	-0.25	0.08	-	-	-	-	-
07	13	Hartogsveld, van Ruiter	Hartogsveld et al., 2020	Group difference on responses to devalued outcomes	Accuracy	-	-	-	-	-	-	-	6.26	1	63	0.62	-0.61	0.07	-	2.50	-	-	-
07	14	Hartogsveld, van Ruiter	Hartogsveld et al., 2020	Group difference on responses to valuable outcomes	Accuracy	-	-	-	-	-	-	-	5.19	1	63	0.57	-0.56	0.07	-	2.28	-	-	-
08	15	Kamp et al.	Kamp et al., 2019	Group difference stress vs control	Accuracy	-	-	-	-	-	-	-	-	1	35	0.73	-0.71	0.12	-	2.22	-	-	-
09	16	Lemmens et al.	Lemmens et al., 2022	Reversal Index (Avoidance)	RI	-	-	-	-	-	-	-	0.21	1	59	0.12	-0.12	0.07	-	0.46	-	-	-
09	17	Lemmens et al.	Lemmens et al., 2022	Groupeffect	RI	-	-	-	-	-	-	-	5.14	1	59	0.58	-0.57	0.07	-	2.27	-	-	-
10	18	Lighthall et al.	Lighthall et al., 2013	Group (Stress vs Control) on Choose	Accuracy	-	-	-	-	-	-	-	10.23	1	88	0.65	-0.65	0.04	-	3.20	-	-	-
10	19	Lighthall et al.	Lighthall et al., 2013	Group (Stress vs Control) on Avoid	Accuracy	-	-	-	-	-	-	-	1.63	1	88	0.26	-0.26	0.04	-	1.28	-	-	-
10	20	Lighthall et al.	Lighthall et al., 2013	Main effect of Stress on sensitivity to recent feedback	Accuracy	-	-	-	-	-	-	-	17.70	1	88	0.86	-0.85	0.05	-	4.21	-	-	-
11	21	Paul et al.	Paul et al., 2018	Main effect of Group (Stress vs Control) on accuracy	Accuracy	-	-	-	-	-	-	-	0.20	1	30	0.16	-0.15	0.13	-	0.45	-	-	-
12	22	Pritchard et al.	Pritchard et al., 2018	Group difference in extinction after devaluation	valued vs devalued	-	-	-	-	-	-	-	5.56	1	68	0.56	-0.56	0.06	-	2.36	-	-	-
12	23	Pritchard et al.	Pritchard et al., 2018	Group difference in reacquisition after devaluation	valued vs devalued	-	-	-	-	-	-	-	2.27	1	68	0.36	-0.36	0.06	-	1.51	-	-	-
13	24	Quaedflieg et al.	Quaedflieg et al., 2019	Condition (Stress vs Control) predicting goal-directed score	valued vs devalued	-	-	-	-	-	-	-	-	1	87	0.94	-0.93	0.05	-	4.56	-	-	-
14	25	Raio et al.	Raio et al., 2020	stage 1 × stress	% stay	-	-	-	-	-	-	-	-	-	-	0.01	-0.01	0.09	-	-	0.02	0.17	-
14	26	Raio et al.	Raio et al., 2020	stage 2 × stress	% stay	-	-	-	-	-	-	-	-	-	-	0.10	-0.10	0.09	-	-	0.18	0.08	-
15	27	Smeets et al.	Smeets et al., 2019	Main effect stress vs. Control	valued vs devalued	-	-	-	-	-	-	-	2.05	1	69	0.34	-0.34	0.06	-	1.43	-	-	-
16	28	Vogel & Schwabe	Vogel & Schwabe, 2018	Explicit knowledge	Accuracy	-	-	-	-	-	-	-	-	-	-	0.54	0.54	0.07	-	2.07	-	-	-
17	29	Wieland et al.	Wieland et al., 2023	Main effect stress vs. Control	Accuracy	-	-	-	-	-	-	-	-	-	-	0.07	-0.06	0.15	-	-	0.12	0.05	-

b)

studyID	effectsizeID	Study_plot	Comparison_specific	statistics for comparison	F	df1	df2	n2p	Cohens_f	r	fishers_z	var_z	t
01	01	Bogdan et al., 2011	Stress vs. Control	Response bias	5.88	1	74	0.07	0.28	-0.03	-0.59	0.01	-
02	02	Carvalheiro et al., 2021	reward/gain learning	Accuracy	12.87	1	19755	0.00	0.03	0.00	-0.16	0.02	-
02	03	Carvalheiro et al., 2021	punishment/loss learning	Accuracy	0.14	1	19755	0.00	0.00	0.00	-0.05	0.02	-
03	04	Carvalheiro et al., 2022	stress vs. Control gains	Accuracy	20.23	1	4400	0.00	0.07	0.00	-0.27	0.05	-
03	05	Carvalheiro et al., 2022	stress vs. Control losses	Accuracy	0.32	1	4400	0.00	0.01	0.00	-0.09	0.05	-
04	06	Ehlers & Todd, 2017	stress vs. Control	handgrips reaching 50% grip strength	7.34	1	100	0.07	0.27	-0.03	-0.58	0.01	2.71
04	07	Ehlers & Todd, 2017	stress vs. Control - high reward	handgrips reaching 50% grip strength	8.52	1	110	0.07	0.28	-0.02	-0.59	0.01	2.92
05	08	Goldfarb et al., 2017	Group×Cue×Day effect	Context vs. SR Memory	5.67	1	56	0.09	0.32	-0.04	-0.64	0.02	-
06	09	Goldfarb et al., 2018	Main effect Stress	Accuracy	2.05	1	28	0.07	0.27	-0.05	-0.58	0.04	-
06	10	Goldfarb et al., 2018	Main effect Stress	Accuracy	4.33	1	27	0.14	0.40	-0.07	-0.75	0.04	-
07	11	Hartogsveld et al., 2020	Group × Value (devalued vs valuable)	Accuracy	7.68	1	63	0.11	0.35	-0.04	-0.68	0.02	-
08	12	Kruse et al., 2018	CS × Time × Stress on reaction time	RT	6.53	1	53	0.11	0.35	-0.04	-0.68	0.02	-
08	13	Kruse et al., 2018	Main effect Stress on reaction time	RT	5.31	1	53	0.09	0.32	-0.04	-0.64	0.02	-
09	14	Larrosa et al., 2017	Group × Trial (training tail vs testing)	Accuracy	6.60	1	16	0.29	0.64	-0.11	-1.10	0.07	-
10	15	Lemmens et al., 2022	Group x stimulus in reversal	Reversal index	5.14	1	59	0.08	0.30	-0.04	-0.61	0.02	-
11	16	Molins et al., 2025	Group (Stress vs Control)	IG Total score	2.19	1	72	0.03	0.17	-0.02	-0.44	0.01	-
12	17	Paul et al., 2018	Stimulus type (Typical vs Exception) × Group	Accuracy	2.58	1	30	0.08	0.29	-0.05	-0.61	0.03	-
13	18	Paul et al., 2020	Group × Feedback-Delay interaction	Accuracy	5.21	1	36	0.13	0.38	-0.06	-0.72	0.03	-
14	19	Radenbach et al., 2015	Main effect of Stress on stay/switch	Stay probability	5.69	1	31	0.16	0.43	-0.07	-0.78	0.03	-
15	20	Schwabe & Wolf, 2010	Group × Value × Time	valued vs devalued	13.61	1	46	0.23	0.54	-0.06	-0.95	0.02	-
16	21	Smeets et al., 2018	Group x Value	valued vs devalued	2.05	1	69	0.03	0.17	-0.02	-0.44	0.01	-
17	22	van Ruitenbeek et al., 2021	Main effect stress vs. Control	Accuracy	3.91	1	89	0.04	0.21	-0.02	-0.49	0.01	-
17	23	van Ruitenbeek et al., 2021	Group x devalued	Accuracy	5.45	1	91	0.06	0.24	-0.02	-0.54	0.01	-
18	24	Voulgaropoulou et al., 2022	Group x Trial Type (RL vs. EL)	Accuracy	6.53	1	78	0.08	0.29	-0.03	-0.60	0.01	-
19	25	Walter et al., 2021	Main effect stress vs. Control	Accuracy	0.47	1	164	0.00	0.05	0.00	-0.24	0.02	-

Additional studies included in the sensitivity analysis

studyID	effectsizeID	Study_plot	Country	Group_Allocation	Age_N1_Stress	Age_N2_Control	Gender	N1	NC	N	Stressor		
Primary analysis													
04	10	de Berker et al., 2016	United Kingdom	R	NA	NA	m/f	25	30	55	SECPT		
04	11	de Berker et al., 2016	United Kingdom	R	NA	NA	m/f	25	30	55	SECPT		
Secundary analysis													
02	02	Carvalheiro et al., 2021	Portugal	NA	21.90	21.90	m	NA	NA	62.00	auditory stimulus		
02	03	Carvalheiro et al., 2021	Portugal	NA	21.90	21.90	m	NA	NA	62.00	auditory stimulus		
03	04	Carvalheiro et al., 2022	Portugal	NA	23.00	23.00	m	NA	NA	23.00	auditory stimulus		
03	05	Carvalheiro et al., 2022	Portugal	NA	23.00	23.00	m	NA	NA	23.00	auditory stimulus		
04	06	Dunsmoor et al., 2017	USA	ns	22.20	23.60	m/f	17	19	36	CPT		
16	21	Schwabe & Wolf, 2009	Germany	ns	23.60	23.60	m/f	29	31	60	SECPT		
18	23	Schwabe & Wolf, 2011	Germany	R	ns	ns	m/f	32	32	64	SECPT		
18	24	Schwabe & Wolf, 2011	Germany	R	ns	ns	m/f	32	32	64	SECPT		
22	29	Walter et al., 2021	Australia	NR	22.43	22.43	m/f	NA	NA	57.00	TSST		
23	30	Zhang et al., 2020	China	ns	ns	ns	m/f	ns	ns	24	Threat of shock		

studyID	effectsizeID	Study_plot	Learning_Type	Comparison_specific	stress_time_concurrent	F	df1	df2	n2p	Cohens_f	var_f	r	g
Primary analysis													
04	10	de Berker et al., 2016	Instrumental/RL	NoGo to Win (Stress vs. Control)	% of go choices	-	-	-	-	-	-	-	1.00
04	11	de Berker et al., 2016	Instrumental/RL	NoGo to Avoid Losing (Stress vs. Control)	% of go choices	-	-	-	-	-	-	-	1.00
Secundary analysis													
02	02	Carvalheiro et al., 2021	Instrumental	reward/gain learning	Accuracy	12.87	1	19755	0.00	0.03	0.02	0.00	
02	03	Carvalheiro et al., 2021	Instrumental	punishment/loss learning	Accuracy	0.14	1	19755	0.00	0.00	0.02	0.00	
03	04	Carvalheiro et al., 2022	RL	stress vs. Control gains	Accuracy	20.23	1	4400	0.00	0.07	0.05	0.00	
03	05	Carvalheiro et al., 2022	RL	stress vs. Control losses	Accuracy	0.32	1	4400	0.00	0.01	0.05	0.00	
04	06	Dunsmoor et al., 2017	Pavlovian	stimulus × group interaction	shock expectancy	-	-	-	-	-	-	0.00	
16	21	Schwabe & Wolf, 2009	Instrumental	Learningcurve group	% high probability choices	-	-	-	-	-	-	0.00	
18	23	Schwabe & Wolf, 2011	Instrumental	Learningcurve group	% high probability choices	-	-	-	-	-	-	0.00	
18	24	Schwabe & Wolf, 2011	Instrumental	block type (last block) x group interactions	% high probability choices	-	-	-	-	-	-	0.00	
22	29	Walter et al., 2021	RL	Main effect stress vs. Control	Accuracy	0.47	1	164	0.00	0.05	0.02	0.00	
23	30	Zhang et al., 2020	RL	Main effect stress vs. Control	Accuracy	-	-	-	-	-	-	0.00	