

University of Groningen

The rodent object-in-context task

Sep, Milou S. C.; Vellinga, Marijn; Sarabdjitsingh, R. Angela; Joels, Marian

Published in:
 PLoS ONE

DOI:
[10.1371/journal.pone.0249102](https://doi.org/10.1371/journal.pone.0249102)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Sep, M. S. C., Vellinga, M., Sarabdjitsingh, R. A., & Joels, M. (2021). The rodent object-in-context task: A systematic review and meta-analysis of important variables. *PLoS ONE*, *16*(7), 1-19. [0249102].
<https://doi.org/10.1371/journal.pone.0249102>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

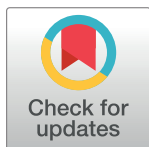
RESEARCH ARTICLE

The rodent object-in-context task: A systematic review and meta-analysis of important variables

Milou S. C. Sep^{1,2}*, Marijn Vellinga¹, R. Angela Sarabdjitsingh¹, Marian Joëls^{1,3}

1 Department of Translational Neuroscience, UMC Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands, **2** Brain Research and Innovation Centre, Ministry of Defence, Utrecht, The Netherlands, **3** University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

* These authors contributed equally to this work.

* m.s.c.sep@umcutrecht.nl**OPEN ACCESS**

Citation: Sep MSC, Vellinga M, Sarabdjitsingh RA, Joëls M (2021) The rodent object-in-context task: A systematic review and meta-analysis of important variables. *PLoS ONE* 16(7): e0249102. <https://doi.org/10.1371/journal.pone.0249102>

Editor: Patrizia Campolongo, Università degli Studi di Roma La Sapienza, ITALY

Received: March 9, 2021

Accepted: June 24, 2021

Published: July 16, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0249102>

Copyright: © 2021 Sep et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data and code files are available from the Open Science Framework (OSF) database (<https://osf.io/gy2mc/>).

Funding: This study was supported by ZonMW grant 'Meer Kennis met Minder Dieren' module

Abstract

Environmental information plays an important role in remembering events. Information about stable aspects of the environment (here referred to as 'context') and the event are combined by the hippocampal system and stored as context-dependent memory. In rodents (such as rats and mice), context-dependent memory is often investigated with the object-in-context task. However, the implementation and interpretation of this task varies considerably across studies. This variation hampers the comparison between studies and—for those who design a new experiment or carry out pilot experiments—the estimation of whether observed behavior is within the expected range. Also, it is currently unclear which of the variables critically influence the outcome of the task. To address these issues, we carried out a preregistered systematic review (PROSPERO CRD42020191340) and provide an up-to-date overview of the animal-, task-, and protocol-related variations in the object-in-context task for rodents. Using a data-driven explorative meta-analysis we next identified critical factors influencing the outcome of this task, such as sex, testbox size and the delay between the learning trials. Based on these observations we provide recommendations on sex, strain, prior arousal, context (size, walls, shape, etc.) and timing (habituation, learning, and memory phase) to create more consensus in the set-up, procedure, and interpretation of the object-in-context task for rodents. This could contribute to a more robust and evidence-based design in future animal experiments.

1. Introduction

Context is defined as a set of independent features that can be observed by an individual and which are stable aspects of the environment [1, 2]. Context-dependent, or contextual, memory is a specific type of episodic memory in which information of events is stored in combination with contextual features [3, 4]. Being able to remember an event with the corresponding contextual information is highly adaptive, since it enables an individual to adjust behavior and respond adequately when encountering a similar event again in a comparable context [2, 5].

(project #114024150) and the Dutch Ministry of Defense. MS is supported by a personal grant which is part of the Graduate Program (project #022.003.003) of The Netherlands Organization of Scientific Research NWO. MS and MJ were supported by the Consortium on Individual Development (CID), which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and Netherlands Organization for Scientific Research (project #024.001.003). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Conversely, generalizing the response to different contexts may be maladaptive and may contribute to the etiology of psychopathologies, e.g. posttraumatic stress disorder [2], panic disorders [6], phobias [5] or Alzheimer's disease [7].

Context-dependent memory with neutral valence—in contrast to contextual classical or operant conditioning [8]—is in humans often experimentally investigated via the combined presentation of items—like faces [9–12], everyday objects [11–14], or words [15, 16]—and contexts—like scene pictures [9–12, 15, 16], words [14] or sounds [13]—in a computer task [17].

In rodents, the most widely used task to assess neutral context-dependent memory uses physical objects instead of virtual items. This task is commonly known as the object-in-context task (OIC) [18] and relies on the hippocampal system and associated cortical regions [1, 19], comparable to human context-dependent memory [14, 17, 20]. In the OIC test the hippocampal system is needed to establish the link between the object's features and the contextual features; that is, without the hippocampal system an animal is able to remember the object in the sense of familiarity, but not remember the context in which it was encountered [19]. Performance in the OIC task is often used as a behavioral measure of hippocampal function [21]. Moreover, the task is frequently applied to probe context-dependent memory in disease or adversity models, e.g. in animal models of Alzheimer's disease [22], substance (ab)use [23] or early life stress [24].

The OIC task is based on the rodent's spontaneous exploration of objects, which requires (almost) no training [25]. As summarized in Fig 1, the task typically consists of two consecutive phases. In the sample phase the rat freely explores an environment—context A—in which two similar objects are located. This is followed by a second trial in which the rodent is placed in a second environment—context B—with different contextual features and a different set of similar objects. After a certain delay (learning-memory retention time) the test phase is

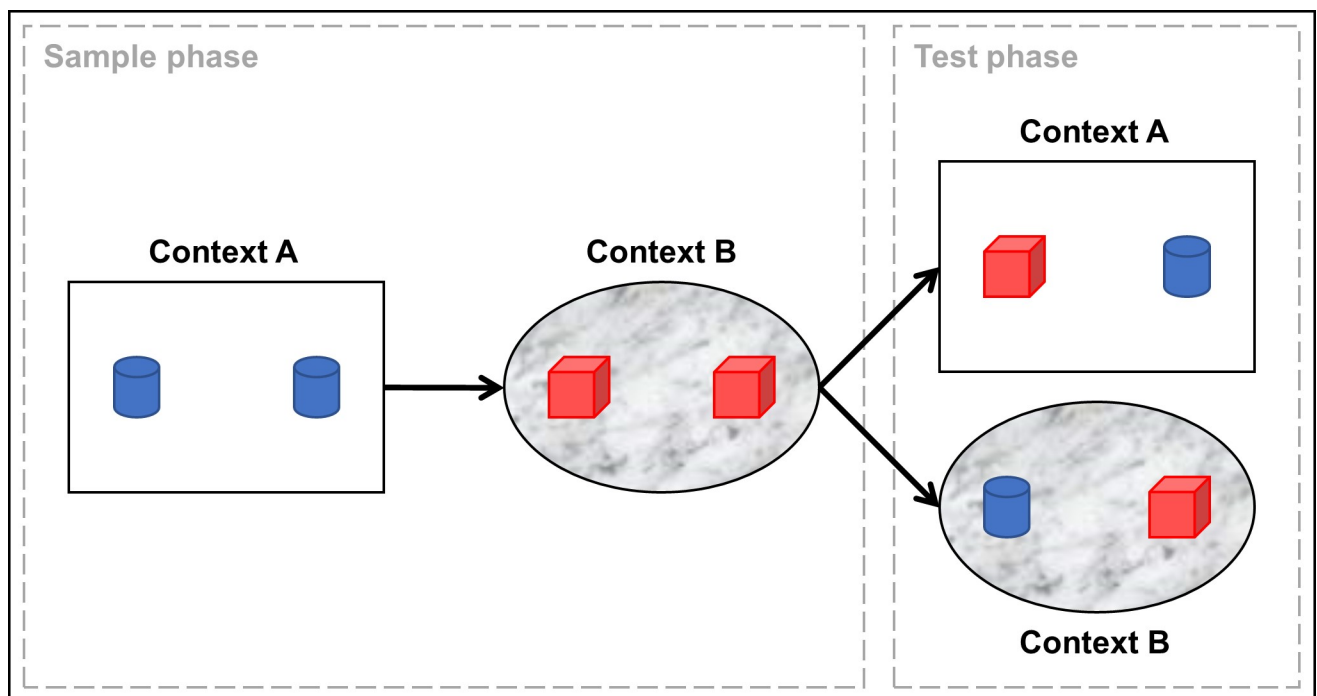


Fig 1. Schematic overview of the rodent object-in-context task (OIC). During the sample phase (learning) an animal encounters a unique set of two objects in context A and next a different set in B. In the test phase (memory) the animal is exposed to either context A or B, with one object of each unique set.

<https://doi.org/10.1371/journal.pone.0249102.g001>

conducted, in which the rat is placed in either context A or B, but this time with one object previously encountered in context A and one object previously encountered in context B. As a result, one object is new in that environment (i.e. novel), while the other has been encountered before in the same environment (i.e. familiar). In the OIC task, contextual information is needed to discriminate between objects. According to the so-called novelty preference paradigm [26] it is assumed that rodents have a preference for novel over familiar objects. Animals will explore the novel object more often if they remember the object-context combinations from the sample phase, reflecting their context-dependent memory. The main outcome measure of the OIC task is the discrimination ratio (DR), which is calculated for each animal. This is based on the time the animal spends exploring the novel object in relation to the time spent exploring the familiar object and the total exploration time.

Despite these general principles of the task, there is a great deal of variation possible in the set-up, procedures and interpretation. Variations in set-up and procedure might have consequences for the animals' behavioral performance and hinder direct comparison between studies. To chart these variations and particularly to identify critical factors determining experimental outcome, we performed a systematic literature review of studies that employed the OIC task in control animals, i.e. naive, sham-operated or saline-injected rodents. As a follow-up, we determined the average DR for control animals using a meta-analytical approach. Next, a data-driven exploration was used to identify which variations affect the behavioral outcomes of this task. Considerable variation in OIC implementation among published studies is expected, and we expect that (some of) these variations affect animals' performance. In the Discussion, we integrated the observations into methodological recommendations for future studies using the OIC task and provide a critical reflection on the (novelty preference) assumptions of the OIC.

2. Methods

This study was performed and reported in accordance with the SYRCL [27, 28], PRISMA [29] and ARRIVE guidelines [30], and preregistered in PROSPERO (CRD42020191340) [31]. The PRISMA checklist is provided in A10 Appendix in [S1 File](#). The documents, datasets and code used during literature search, screening, data extraction, and meta-analyses are available via Open Science Framework (OSF; <https://osf.io/gy2mc/>). The collected dataset allows for the development of web-based explorative tools, that can become available on the OSF webpage.

2.1. Search strategy

A comprehensive literature search was performed in the electronic database PubMed. The search string contained search terms for '*learning and memory*', '*rodents*', '*object-in-context task*' and terms to exclude *meta-analyses* and *systematic reviews* (the complete search filter is provided in A1 Appendix in [S1 File](#)). The final search was performed on 25th of May 2020. References of included publications were checked for eligibility (snowballing).

2.2. Screening

The retrieved articles were screened based on *a priori* defined inclusion and exclusion criteria. Inclusion criteria for the systematic review were: (1) original article in the English language, (2) OIC task, (3) rodents, (4) control animals (no treatment, saline injection, or sham operation). Studies had to comply with the inclusion criteria above and report the sample size and (the data to calculate) the DR-with the corresponding standard error- to be included in the meta-analysis. Exclusion criteria were: 1) no primary literature in English; 2) other measures of context-dependent memory, like modified versions of the OIC task (e.g. combinations of

context and location measures, context-dependent memory based on odors, classical fear conditioning or operant conditioning paradigms); 3) non-rodent species, 4) no control group tested. Note, genetic modification was not an exclusion criterium. Screening was performed by MV and MS. If information in the title and abstract was insufficient to determine eligibility, full-text articles were checked.

2.3. Data extraction and study quality assessment

Extracted data was *a priori* defined. The complete data-extraction codebook is provided in A2 Appendix in [S1 File](#), A1 Table in [S1 File](#), and includes: 1) Publication details (authors, year of publication, etc.); 2) methodological details, for instance context-dependent differences between boxes, object characteristics, habituation protocol, trail duration, retention time between phases, order in which contexts are encountered, behavioral scoring, etc.; 3) animal characteristics such as species, strain, sex, age, previous use in experiments, type of control group, housing conditions (e.g. day-night cycle, group-housing) etc. In addition, sample size and time spent exploring the objects to calculate the DR (with standard error) were extracted for studies included in the meta-analysis. Note, the mean sample size was calculated if only a range was provided. Plot Digitizer [32] was used to extract visually presented data from graphs. Authors were not contacted for missing or additional data, missing values were included in dataset and further processed as described in section 2.4.2. Initial data-extraction was performed by MV, 20% of the studies were independently checked by MS.

As multiple formulas to calculate DR for object recognition are described in literature [33], all extracted DRs were transformed to center around 0 (i.e. $DR = (\text{novel} - \text{familiar}) / (\text{novel} + \text{familiar})$). The corresponding standard errors of transformed DRs were recalculated accordingly. For transformation formulas, see A3 Appendix in [S1 File](#).

Study quality and the risk of bias were assessed with SYRCLE's risk of bias tool [34] by MV. Unreported details were scored as an *unclear* risk of bias. MS independently scored 10% of the studies, discrepancies were discussed until consensus was reached.

2.4. Statistical analysis

Statistical analyses were based on earlier work of our group [35] and performed with $\alpha = .05$ in R version 4.0.3 [36], with the use of packages *dplyr* [37], *osfr* [38], *metafor* [39], *metaforest* [40], *caret* [41] and *ggplot2* [42].

2.4.1 Random-effects meta-analysis: Estimation of overall effect. The (transformed) observed mean DR was used as effect size, i.e. raw mean. Sampling variance was calculated from the (transformed)observed standard error and sample size. As heterogeneity in the data was expected, the overall effect size (i.e. overall mean DR) was estimated with a nested random effects model with restricted maximum likelihood estimation [43]. The estimation was nested within *articles* and *experiments*. Heterogeneity was assessed with Cochrane Q-test [39] and the I^2 -statistic (low: 25%, moderate: 50%, high: 75% [44]). Robustness of the estimated effect was evaluated via *Rosenthal's fail-safe N* [45] and trim-and-fill analyses [46]. Moreover, funnel plot asymmetry -as index for publication bias- was tested via *Egger's regression* [47] and *Begg's test* [48]. Sensitivity analyses were performed to evaluate if the estimated effect was influenced by 1) study quality and/or 2) influential cases or outliers [49]. To evaluate the influence of study quality, the scores on SYRCLE's risk of bias tool for randomization (0–5), blinding (0–2) and reporting (0–2) were combined into a summary quality score (formula in the A4 Appendix in [S1 File](#)).

2.4.2 Random forest for meta-analysis: Exploration of heterogeneity. To explore the source heterogeneity (variation between studies), a random forest-based meta-analysis was

performed using *MetaForest* [40]. This data-driven approach allows to rank potential moderators of the overall effect, based on variable importance in the random forest.

Variables with more than 1/3 missing values, indicated in A2 and A7 Appendices in [S1 File](#), were excluded from the random forest analysis. The missing values in other variables were replaced by median value (for continuous variables) or most prevalent category (for categorical variables).

As certain variables are part of one underlying factor, four summary scores were added to the analysis (formulas are provided in the A5 Appendix in [S1 File](#)): one context difference score and three arousal scores: 1) arousal prior to OIC, 2) arousal related to OIC habituation procedures and 3) their combination. Of note, no cumulative object difference score was created for object material and size, as object size contained more than 1/3 missing values and was excluded from analyses (note, the object material variable was included in the analyses).

The random forest-based meta-analysis (600 trees) was tuned (optimal parameters for minimal RMSE: uniform weighting, 4 candidate moderators at each split, and a minimum node size of 2) and 9-fold cross-validated.

To follow-up the most important moderators, i.e. the upper 50% based on random forest variable importance, partial dependence (PD) plots were used to explore the relations between moderator and DR. PD plots show the predicted effect at different levels of a particular moderator, if all other moderators are kept constant [40, 50]. In addition, weighted scatter plots were created to inspect the distribution of the raw data per moderator.

3. Results

3.1. Study selection and characteristics

After screening 254 unique studies (one publication was identified via snowballing), 41 articles were included in the systematic review. Four of these did not report on sample size, standard error or standard deviation and had to be excluded from the meta-analyses. In total 37 papers with 857 unique animals were included in the analyses. The flowchart is shown in [Fig 2](#) and the study characteristics are provided in A6 Appendix in [S1 File](#) (A2 Table in [S1 File](#)).

The results of the systematic review are shown in A7 Appendix in [S1 File](#) (A3 Table in [S1 File](#)), and reveal extensive variation in animal characteristics, set-up and task procedures' related factors. Although all rodent species were eligible, the screening process only returned rat and mice papers. Study's quality is shown per item in [Fig 3](#) and the cumulative study scores are shown in A8 Appendix in [S1 File](#) (A1 Fig in [S1 File](#)). Many studies did not report on all potential risks of bias ([Fig 3](#)): only 6 out of 41 articles (~15%) reported that 4 or more -out of SYRACLE's 9 [34]- measures were taken on randomization, blinding and / or reporting to reduce the risk of bias (A1 Fig in [S1 File](#)).

3.2. Meta-analysis of Discrimination Ratio (DR)

We next performed a meta-analysis on the available data (forest plot in [Fig 4](#)). This revealed a high degree of heterogeneity in the multilevel model ($Q(80) = 406.307$, $p < .0001$; $I^2 = 82.994$); 77% due to variance between studies and 6% due to variance within studies. The overall estimated DR was significantly different from 0 (mean DR = 0.2579, SE = 0.0266, 95% CI = 0.2057–0.3101, $z = 9.6879$, $p < .0001$), indicating that animals discriminated significantly between the in- and out-of-context objects. The standard deviation of the estimated overall DR is 0.7785 ($SD = SE * \sqrt{n}$ unique animals in meta-analysis [51] = $0.0266 * \sqrt{856.5}$) and the estimated effect size of the difference between the estimated overall DR and 0 is small to

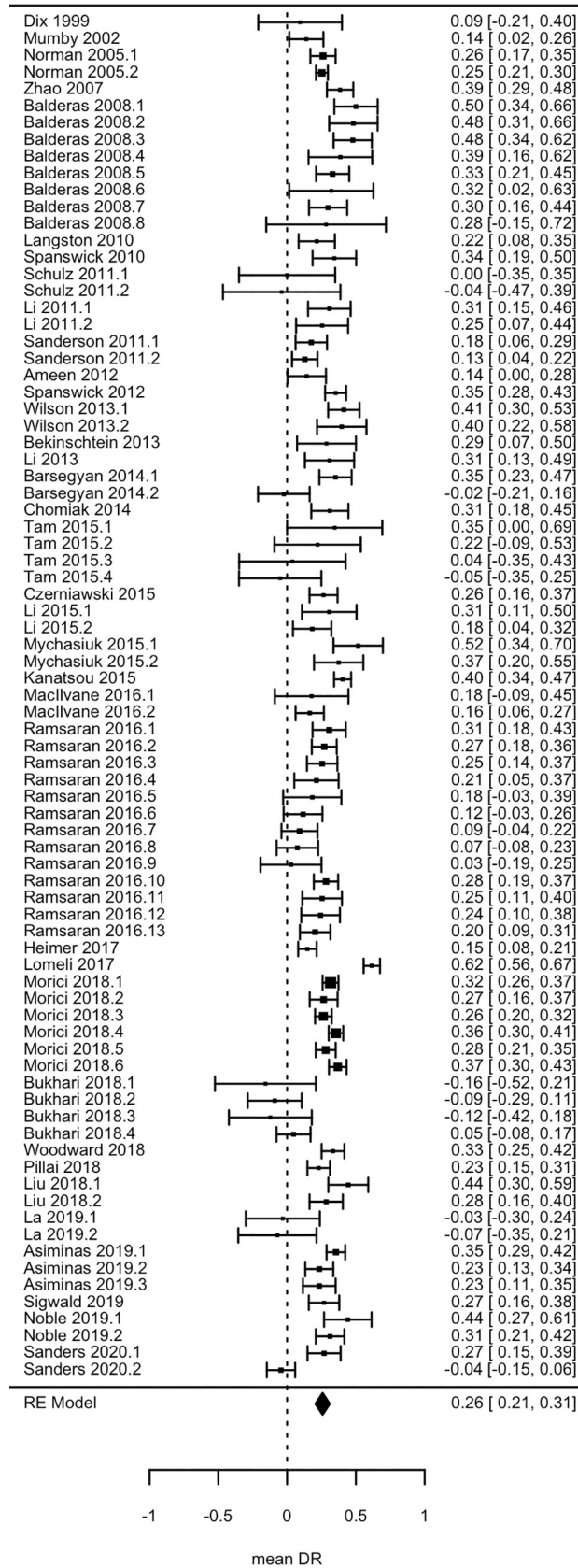


Fig 4. Forest plot. Forest plot visualizing the Discrimination Ratio (DR) per experimental group, per study; and the overall mean DR with 95% Confidence Interval. Studies are presented by publication year in ascending order.

<https://doi.org/10.1371/journal.pone.0249102.g004>

3.3. Robustness of the estimated DR: Sensitivity analyses and publication bias

The presence of publications bias is suggested by qualitative examination of funnel plot asymmetry (Fig 5), and was confirmed by Egger's regression ($z = -3.492$, $p < .001$) and Begg's test ($z = -3.4925$, $p = .0005$). However, file drawer analyses suggested that the influence of publication bias was limited: 12 (SE = 5.8941) studies were missing on the right side, according to trim-and-fill analysis, and 43154 studies would be needed to nullify the estimated effect, according to Rosenthal's fail-safe N analysis.

Furthermore, sensitivity analyses revealed that 1) study quality -assessed with SYRACLE's RoB tool- did not moderate the estimated overall DR (QM(1) = 1.9511, $p = 0.1625$); 2) there were no potential influential cases; and 3) there were four potential outliers, though without substantial impact on the results, i.e. their exclusion led to a similar estimated overall DR: mean DR [95%CI] = 0.2599 [0.2165,0.3033], $z = 11.7398$, $p < .0001$).

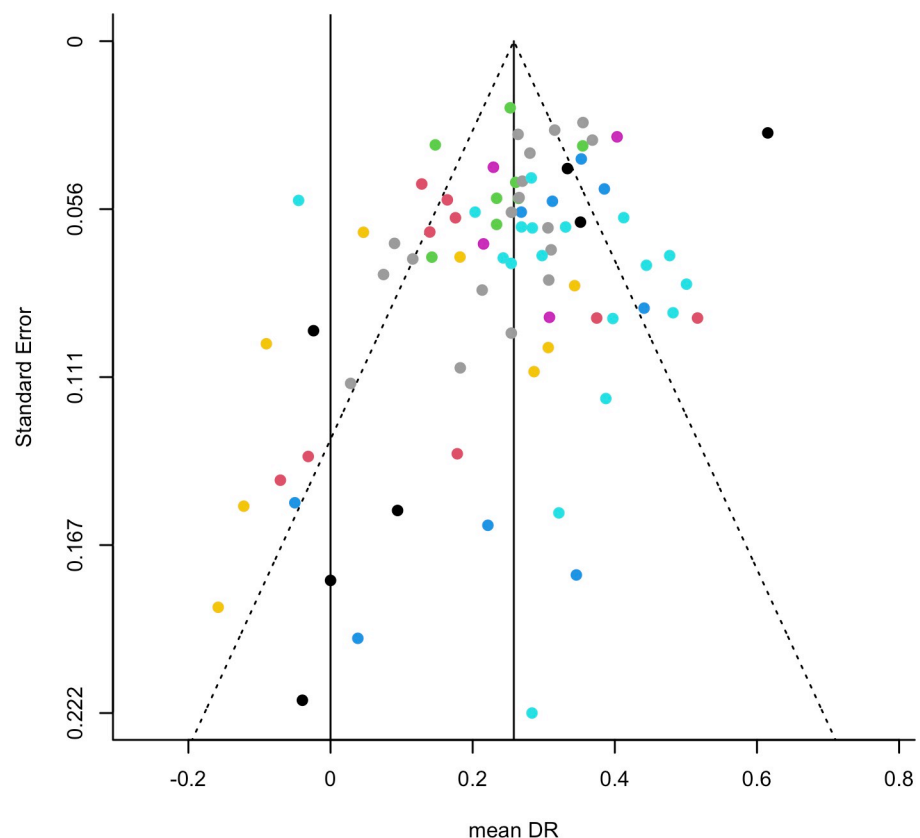


Fig 5. Funnel plot. Funnel plot showing the DR of individual control groups on the x-axis against their standard errors (i.e., the square root of the sampling variances) on the y-axis. Vertical reference lines indicate the 0, i.e. no context-dependent memory; and the estimated overall mean DR based on the model. Colors indicate unique studies (PMID).

<https://doi.org/10.1371/journal.pone.0249102.g005>

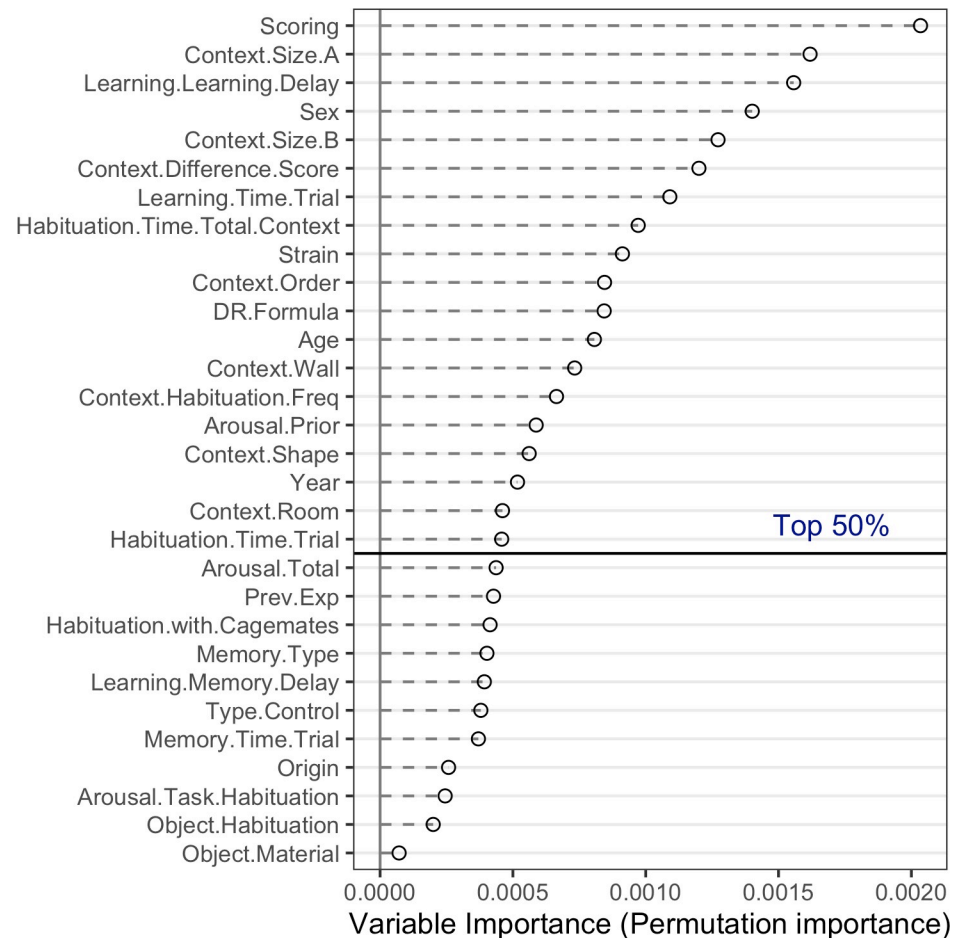


Fig 6. Relative importance of potential moderators based on 'permuted variable importance' in the random forest-based meta-analyses. Moderators above the horizontal line belong to the top 50%. Full definitions of variables and sum scores can be found in A2 and A5 Appendices in [S1 File](#) respectively.

<https://doi.org/10.1371/journal.pone.0249102.g006>

3.4. Exploration of moderators: Random-Forest variable importance

The data-driven exploratory random forest-based meta-analysis provided insight in the sources of heterogeneity in the sample. Our 9-fold cross-validated random forest model showed good convergence (see A9 Appendix in [S1 File](#), A2 Fig in [S1 File](#)) and the included moderators accounted for 37.5% of the variance ($Rcv^2[SD] = 0.375[0.215]$). The ranking of potential moderators—based on permuted random forest variable importance—is shown in [Fig 6](#).

The upper 50% of the most important variables included four animal-related factors (Sex, Strain, Age, Arousal.Prior), six set-up related factors (Context.Size.A, Context.Size.B, Context.Difference.Score, Context.Wall, Context.Shape, Context.Room), and seven procedure-related factors (Scoring, Learning.Learning.Delay, Learning.Time.Trial, Habituation.Time.Total.Context, Context.Order, Context.Habituation.Freq, Habituation.Time.Trial) as well as two other factors (DR.formula and Year). Note, precise variable and sum score definitions are provided in A2 and A5 Appendices in [S1 File](#) respectively.

Exploratory partial dependence plots show the predicted relations of these top 50% selected moderators and the DR in the OIC task, for rats and mice separately ([Fig 7](#)). Considering the top-5 most influential factors in both rats and mice, it was evident that DR was higher in males

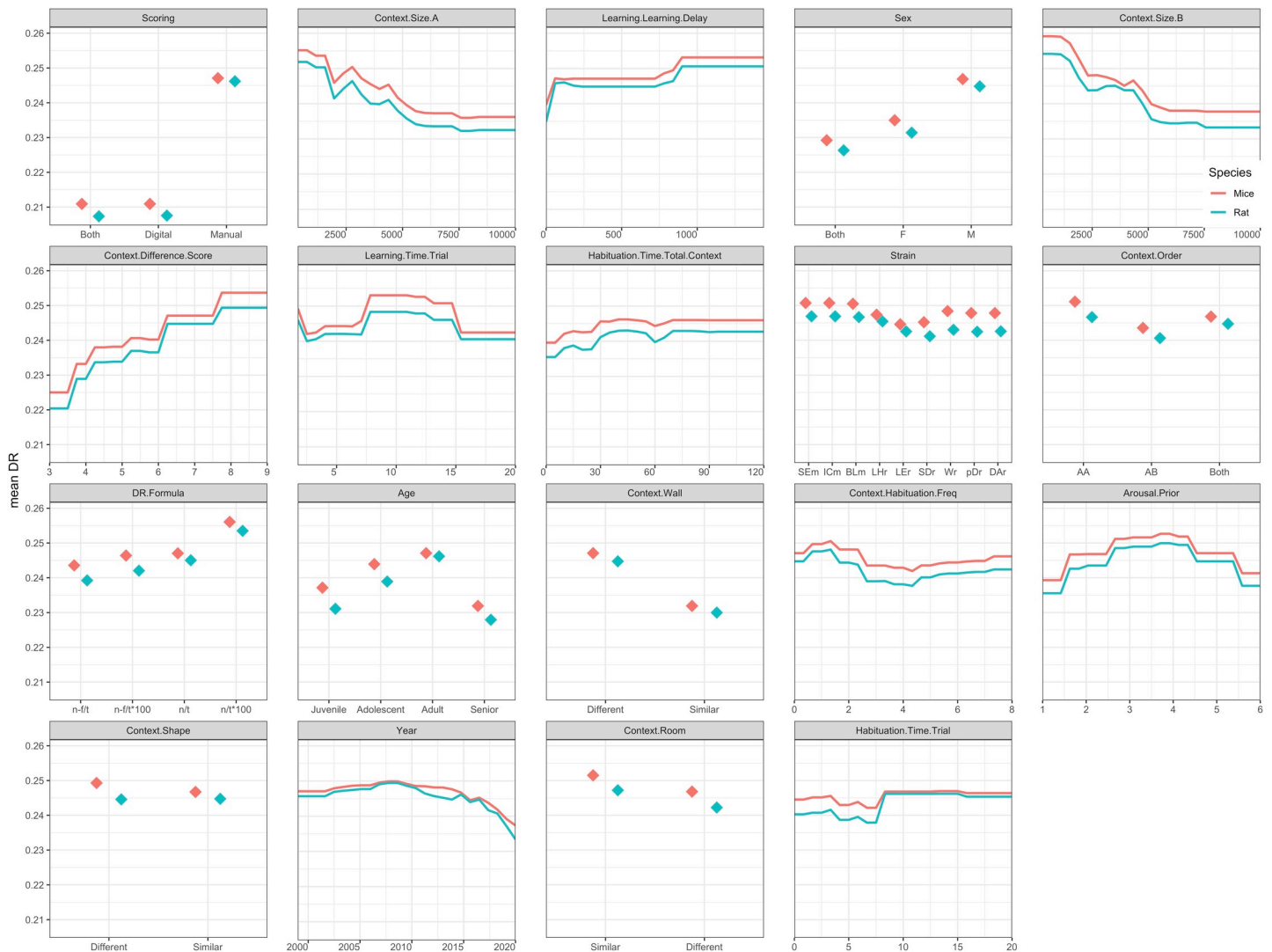


Fig 7. Partial dependence plots showing the predicted relation between DR and the upper half most important variables in the random forest-based meta-analysis, broken down by species: Mice in red vs. rat in blue. The y-axis shows the predicted DR and x-axis shows the values of the variable that is named above the graph (in gray). The values of context.size A and B are shown in in cm^2 and the values of time variables (learning, learning.delay, learning.time.trial, habituation.time.total.context, habituation.time.trial) are shown in minutes. Higher context.difference.score values indicate more different contexts and higher arousal.prior values indicate more arousal prior to the experiment. Strains values are abbreviations: TMm (Tg(Sim1cre)KH21Gsat/Mmucd mice), SEm (SEm Sv/Ev mice), ICm (ICR mice), BLm (C57BL/6 mice), LHr (Lister hooded rats), LEr (Long-Evans rats), SDr (Sprague-Dawley rats), Wr (Wistar rats), pDr (pigmented DA strain rats), DAr (Dark Agouti rats). DR formula values are also abbreviations: n-f/t $((T_{\text{novel}} - T_{\text{familiar}}) / (T_{\text{novel}} + T_{\text{familiar}}))$, n-f/t*100 $([(T_{\text{novel}} - T_{\text{familiar}}) / (T_{\text{novel}} + T_{\text{familiar}})] * 100)$, n/t $(T_{\text{novel}}) / (T_{\text{novel}} + T_{\text{familiar}})$, n/t*100 $([(T_{\text{novel}}) / (T_{\text{novel}} + T_{\text{familiar}})] * 100)$. The context.order value AA indicates that memory was tested in the context of the second learning trial, the value AB indicates that memory was tested in the context of the first learning trial, 'both' indicates that AA and AB were randomized. Finally, context.habituation.freq values show the number of visits per context. Full definitions of variables and sum scores can be found in A2 and A5 Appendices in S1 File respectively. Note, predictions based on the random forest-model do not take into account that strains belong to a specific species (either rat or mouse), hence rat and mice predictions were generated of all strains.

<https://doi.org/10.1371/journal.pone.0249102.g007>

(versus females or mixed populations); when animals were tested in relatively small ($< 2500 \text{ cm}^2$) context boxes (A and B); and when the delay between the two learning sessions was not too short ($> 825 \text{ min}$). Also, the method of scoring (manual versus digital) was found to be of influence: manual scoring led to a higher DR.

Weighted scatter plots showing the distribution of the raw DR by the top 50% selected moderators are presented in a Fig 8. From the top-5 most influential factors, the most variation in DR between animals was observed for males (versus females or mixed groups); when larger

context boxes were used (~ 2500 cm²); and when a very brief delay between learning trials was applied. Variation in the DR was also affected by the scoring method (manual vs digital).

4. Discussion

Object-in-context learning is frequently used to probe contextual memory formation and (dorsal) hippocampal function in rodent models of disease or adversity [54]. However,

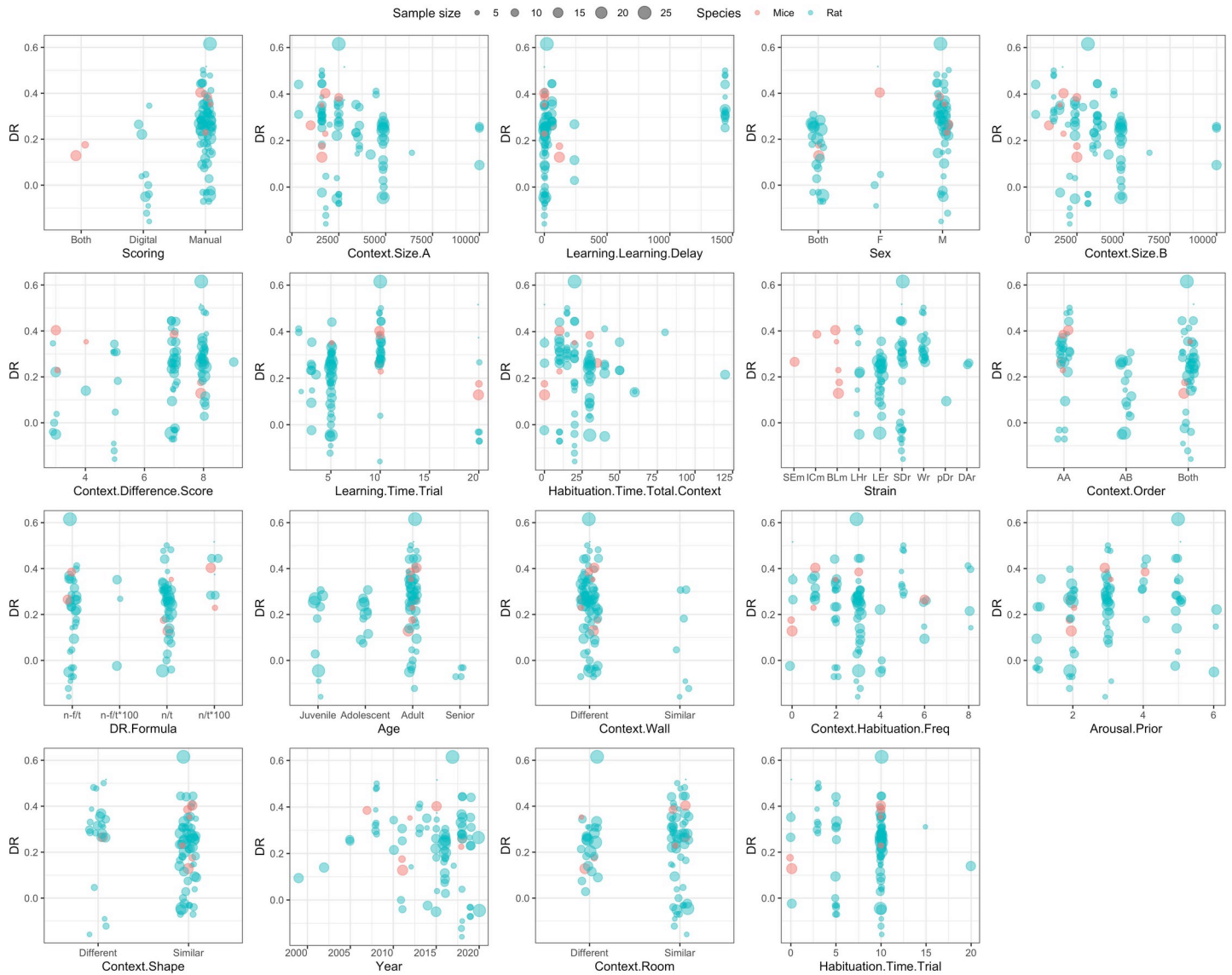


Fig 8. Weighted scatter plots showing the distribution of the raw DR for the 50% most important variables in the random forest-based meta-analysis, broken down by species: Mice in red vs. rat in blue. Dot size indicates the sample size of each observation. Most studies were performed in rats. The y-axis shows the mean DR and x-axis shows the values of the variable that is named below the graph. The values of context.size A and B are shown in cm² and the values of time variables (learning, learning.delay, learning.time.trial, habituation.time.total.context, habituation.time.trial) are shown in minutes. Higher context.difference.score values indicate more different contexts and higher arousal.prior values indicate more arousal prior to the experiment. Strains values are abbreviations: TMm (Tg(Sim1cre)KH21Gsat/Mmucd mice), SEm (SEM Sv/Ev mice), ICm (ICR mice), BLm (C57BL/6 mice), LHR (Lister hooded rats), LER (Long-Evans rats), SDR (Sprague-Dawley rats), Wr (Wistar rats), pDr (pigmented DA strain rats), DAr (Dark Agouti rats). DR.formula values are also abbreviations: n-f/t ((T_{novel}-T_{familiar})/(T_{novel}+T_{familiar})), n-f/t*100 (((T_{novel}-T_{familiar})/(T_{novel}+T_{familiar}))*100), n/t ((T_{novel})/(T_{novel}+T_{familiar})), n/t*100 (((T_{novel})/(T_{novel}+T_{familiar}))*100). The context.order value AA indicates that memory was tested in the context of the second learning trial, the value AB indicates that memory was tested in the context of the first learning trial, 'both' indicates that AA and AB were randomized. Finally, context.habituation.freq values show the number of visits per context. Full definitions of variables and sum scores can be found in A2 and A5 Appendices in S1 File respectively.

<https://doi.org/10.1371/journal.pone.0249102.g008>

variations in animal-, set-up or procedure-related factors seriously hamper comparison between studies. This was confirmed by the current systematic review and subsequent meta-analysis of 37 studies that employed the OIC task in (control) rats and mice. Overall, the DR differed significant from 0, indicating that rodents on average do discriminate between the in- and out-of-context objects, reflecting the formation of context-dependent memory. Yet, this effect was small to medium and accompanied by a large degree of heterogeneity, mostly (77%) due to variance between studies. As expected, a substantial part (37.5%) of the variance could be explained by a set of moderators identified by a random forest approach, with prominent roles for e.g. sex, size of the boxes and delay between the learning trials. In the following sections we will first discuss methodological considerations of our approach and next provide recommendations regarding the set-up, procedure and interpretation of the OIC task in rodents for future users.

4.1. Methodological strengths and limitations

Based on our samples size (37 studies), robust findings can be expected. The quality of meta-analyses, however, depends on the quality of the studies on which these analyses are based. The fact that only ~15% of all studies reported 4 or more out of SYRCLE's 9 measures to reduce risk of bias [34] might suggest poor study quality. However, study quality did not moderate the estimated overall DR. Moreover, as this percentage was influenced by unreported details in many studies, it is probably an underestimation, since practices like randomization or blinding may have been applied but simply not reported. Still, lack of reporting introduces an unclear risk and at least difficulty to estimate the quality of the studies [55]. The problem of insufficient reporting of experimental details and quality measures in pre-clinical studies has been extensively addressed and previously observed in meta-analysis [35, 56–60].

Since our analyses were based on metadata, we cannot exclude that studies were liable to p-hacking practices, such as post-analysis decisions which variables to report on, whether or not to include outliers or stopping data exploration once a significant p -value was reached [61]. In that case one might expect to see publication bias, for which we found only suggestive evidence in the funnel plot. Subsequent sensitivity analyses did not confirm this. Also, potential outliers among the studies did not affect the overall outcome. All in all, assessments of study quality did not indicate severe limitations in the use of the current dataset.

We adopted the state-of-the-art random forest-based meta-analysis *MetaForest* [40] for an unbiased exploration of the sources of heterogeneity in our dataset. This technique is robust to overfitting, can identify non-linear relationships and is valid for meta-analysis with 20 or more studies [62]. Relative variable importance, derived from the random forest model, was used to identify the most important moderators of the DR in control animals. Not all extracted variables could be included in these explorative analyses, due to too much missing values. As we could not judge the potential importance of these excluded variables, the provided overview might be incomplete.

Partial dependance plots were used to visually inspect the marginal effects of these selected moderators on the DR [62]. Although these plots are the most widely used to explore relations in black box models like random forests [50], the visualized relations need to be interpreted with some caution, especially when based on few or unequally distributed raw observations (e.g. *learning.learning.delay*).

4.2. Rethinking the DR definition

The main outcome measure of the OIC task is the discrimination ratio (DR). Two ways have been used to calculate this ratio: Either $DR = T_{\text{novel}} / T_{\text{total}}$ [63] or $DR = (T_{\text{novel}} - T_{\text{familiar}}) / T_{\text{total}}$ [18].

In our analysis, all extracted DRs were recalculated to center around 0 (i.e. $DR = (T_{\text{novel}} - T_{\text{familiar}}) / T_{(\text{novel} + \text{familiar})}$). Interestingly, the average DR was 0.26 and the estimated effect size of its difference to 0 was small to medium. To detect this effect, future studies would require a sample size of 58 animals per group. As such sample sizes are seldomly seen in pre-clinical control groups, future studies could benefit from the use of historical data on the OIC task to increase statistical power with a limited sample size [64]. Since the OIC task is often used to examine shifts in contextual memory in animal models for disease [22, 23] or (early life) adversity [24], such low values in controls might introduce a ‘floor effect’, i.e. a lower DR in the experimental group might remain undetected given the already low DR in the controls.

Of particular interest is the interpretation of the DR. It is generally assumed [26] that rodents have a preference for novel objects over familiar objects and thus will explore the novel object more often if they remember the object-context combinations from the sample phase. However, one could also argue that consistent preference for the *familiar* object in a context may reflect context-dependent memory, assuming a more ‘conservative strategy’ in which e.g. neophobia or lack of boldness outweigh innate curiosity (reviewed in [65]). If so, DRs that differ considerably from 0 (assuming transfer around 0, with 0 indicating performance at chance level) would reflect the formation of a contextual memory, regardless of the sign. To rule out this influence of differences in strategy (the latter being a composite of many underlying factors like curiosity, boldness or neophobia), one would then prefer to use the absolute value of the difference between time spent with the novel and familiar object in a context, as a function of the total object exploration time [66]. The proposed “absolute DR” [66] is preferable when the aim is to measure (experimental influences on) context-dependent memory, regardless of the animal’s strategy. However, when the aim is to examine (experimental influences on) an animal’s strategy, the sign of the DR (centered around 0) holds valuable information.

4.3. Critical factors contributing to the DR in control animals

Based on the unbiased random forest approach, we conclude that many variables affect the DR in the OIC task. Together, these variables explain 37.5% of the variation between and within studies.

Regarding the animal-related factors, it was interesting that very little difference was observed between studies with rats versus mice, although it should be noted that only few studies (<10%) were carried out with mice, so that any conclusions about mice should be made with care. Among the studies with rats, clear strain differences were observed: Sprague Dawley or Long Evans rats displayed much more variation than e.g. Wistars. Interestingly, these strain differences align with earlier observations in another visuo-spatial learning task [67]. In terms of age, highest DR values were found in adulthood, with on average lower values in younger or older animals. This is of interest, since context-dependent memory in humans shows a similar inverted U-shaped age-dependency [68]. Memory context-dependency increases with age in children, as they develop the ability to bind and integrate information [69]. As adults get older, context dependency decreases, which has been linked to age-related reductions in selective attention -leading to hyper-binding of too much contextual details thereby reducing accuracy for the relevant context- [70, 71]; alterations in prefrontal-hippocampal connectivity [72]; and declines in hippocampal neurogenesis [73, 74]. In line with frequently reported sex-dependence of spatial or neutral contextual memory formation across species [75, 76], male animals showed a higher DR than females, but firm conclusions await more studies in females. Performance in the OIC might be enhanced in males compared to females, as the former are more likely to adopt a hippocampus-dependent place-strategy to navigate an environment (i.e.

navigate based on a contextual map) than the latter (who are more likely to use a striatum-dependent landmark strategy) [77]. Such place-strategies could create a stronger representation of the OIC context in the brain, enabling richer context-dependent memories. Finally, an often-neglected factor concerns (potential) exposure of animals to arousal prior to being tested. As with age, an inverted U-shaped dependency was observed for prior arousal, with highest DR values in moderately aroused animals. Inverted U-shaped dose-dependency is a common phenomenon in stress-related influences on memory formation [78]. Of note, it is also known that arousal and stress can affect memory formation and retrieval differently [79]. As a consequence, the time of saline-injection -which can trigger a mild stress response [80]- with respect to the OIC phases (e.g. before sample or test phase; or exactly when relatively to the test phase) could have caused variation in the saline-injected control animals, which was not accounted for in the current meta-analysis. However, type of control group ranked among the less important variables (bottom 50%), suggesting that variation from this source might have limited effects on overall performance.

Next to animal-related factors, factors related to the experimental set-up or learning paradigm also influenced OIC outcome. In general, it transpired that variation in context affected the outcome much more strongly than object (material) variation. Smaller arenas resulted in higher DR values, as did large difference between context A and B. Recency effects play a role in the DR values, as higher DR values were observed when memory was tested in the context that was used in the last (second) learning trial, as opposed to the first learning trial. Interestingly, higher DR values were also observed when the delay between learning trials was longer, while a shorter delay led to more individual variation in the DR.

Finally, one of the factors contributing most clearly to the variation was the way of scoring: DR values were found to be much higher in studies using manual scoring than in those using automated scoring programs. It cannot be excluded that this was somewhat affected by (the absence of) blinding of the experimenter to groups and/or object type (novel vs familiar). Manual scoring without proper blinding could lead to more subjective interpretation of an animal's interaction with one of the objects, whereas automated scoring is generally combined with blinding of the experimenter, which leads to more trustworthy results.

4.4. Recommendations for future studies

Based on the random forest analysis, recommendations for optimizing DR values are summarized in Table 1. We considered two angles in the design: if one is primarily interested in studying context-dependent memory formation, the task should be designed such that the (absolute) DR value is as high as possible, with very little variation between animals in the control group. However, if one is also interested in individual differences in context-dependent memory formation, variation over the entire spectrum of (absolute) DR values is welcome. Of note, the factors indicated in Table 1 are currently based on qualitative rather than quantitative interpretation. The exact degree to which they contribute to the overall outcome of the test was not determined in the current analysis, which is a limitation of the study.

All in all, the current study illustrates that insights from historical datasets can help to interpret data from control animals, which can next be used to increase power of future studies [64]. Performing an unbiased data-driven analysis of metadata may form the basis for more consensus on the set-up, procedure and interpretation of the OIC task for rodents; and hence for recommendations how to design future studies. This may be particularly helpful for those who have never used the task before. But even for more experienced investigators, awareness of factors influencing the dependent variable may help to optimize the experimental design.

Table 1. Recommendations for future animal studies.

DR calculation:	⇒ Aim to study context-dependent memory: absolute DR = $\text{abs}(T_{\text{novel}} - T_{\text{familiar}}) / T_{\text{total}}$ ⇒ Aim to study animals' strategy: the sign (positive vs negative) of DR as calculated with $(T_{\text{novel}} - T_{\text{familiar}}) / T_{\text{total}}$	
Factors:	1) for the highest mean DR (based on PD plots in Fig 7: ⁽¹⁾ Sex; ⁽²⁾ Strain; ⁽³⁾ Age; ⁽⁴⁾ Arousal.Prior; ⁽⁵⁾ context.size.B & context.size.A; ⁽⁶⁾ context.difference.score; ⁽⁷⁾ context.wall; ⁽⁸⁾ context.shape; ⁽⁹⁾ context.room; ⁽¹⁰⁾ Habituation.time.total.context; ⁽¹¹⁾ context.habituation.freq; ⁽¹²⁾ habituation.time.trial; ⁽¹³⁾ learning.learning.delay; ⁽¹⁴⁾ Learning.Time.Trial; ⁽¹⁵⁾ context.order)	2) for the most individual variation in DR (based on WS plots in Fig 8: ⁽¹⁶⁾ Sex; ⁽¹⁷⁾ Strain; ⁽¹⁸⁾ Age; ⁽¹⁹⁾ context.size.B & context.size.A; ⁽²⁰⁾ context.wall; ⁽²¹⁾ context.shape; ⁽²²⁾ context.room; ⁽²³⁾ context.habituation.freq; ⁽²⁴⁾ habituation.time.trial; ⁽²⁵⁾ learning.learning.delay; ⁽²⁶⁾ Learning.Time.Trial; ⁽²⁷⁾ context.order)
Animal related factors	<ul style="list-style-type: none"> • Males ⁽¹⁾ • Not Long-Evans or Sprague-Dawley rats ⁽²⁾ • Adults ⁽³⁾ • Medium levels of arousal prior to testing ⁽⁴⁾ 	<ul style="list-style-type: none"> • Males ⁽¹⁶⁾ • Long-Evans or Sprague-Dawley rats ⁽¹⁷⁾ • Adults ⁽¹⁸⁾
Set-up related factors	<ul style="list-style-type: none"> • Context boxes < 2500 cm² ⁽⁵⁾ • Most different contexts ⁽⁶⁾, especially: <ul style="list-style-type: none"> ◦ Different walls ⁽⁷⁾ ◦ Different shape ⁽⁸⁾ ◦ Different room ⁽⁹⁾ 	<ul style="list-style-type: none"> • Context boxes ~2500 cm² ⁽¹⁹⁾ • Different context walls ⁽²⁰⁾ • Similar context shape ⁽²¹⁾ • Similar context room ⁽²²⁾
Task procedure related factors	<p>Habituation:</p> <ul style="list-style-type: none"> • Total habituation time per context > 30 minutes ⁽¹⁰⁾ • 2 or less habituation trials per context ⁽¹¹⁾ • > 7.5 minutes habituation per trial ⁽¹²⁾ <p>Learning (sample phase):</p> <ul style="list-style-type: none"> • Delay between learning trials > 825 min ⁽¹³⁾ • Time of learning trials between 7.5–15 minutes ⁽¹⁴⁾ <p>Memory (test phase):</p> <ul style="list-style-type: none"> • Testing in last learning context ⁽¹⁵⁾ 	<p>Habituation:</p> <ul style="list-style-type: none"> • 3 habituation trials per context ⁽²³⁾ • 10 minutes habituation per trial ⁽²⁴⁾ <p>Learning (sample phase):</p> <ul style="list-style-type: none"> • Brief delay between learning trials ⁽²⁵⁾ • 5 minutes learning trials ⁽²⁶⁾ <p>Memory (test) phase:</p> <ul style="list-style-type: none"> • Counterbalanced testing in first and last learning context ⁽²⁷⁾

<https://doi.org/10.1371/journal.pone.0249102.t001>

Supporting information

S1 Checklist. Prisma checklist.

(PDF)

S1 File. Appendix.

(PDF)

Acknowledgments

We are very thankful to Valeria Bonapersona for here valuable input on the statistical analyses. We thank Elbert Geuze for the opportunity to conduct this research.

Author Contributions

Conceptualization: Milou S. C. Sep, Marijn Vellinga, R. Angela Sarabdjitsingh, Marian Joëls.

Data curation: Marijn Vellinga.

Formal analysis: Milou S. C. Sep.

Funding acquisition: Milou S. C. Sep, Marian Joëls.

Investigation: Milou S. C. Sep, Marijn Vellinga.

Methodology: Milou S. C. Sep, Marijn Vellinga, R. Angela Sarabdjitsingh.

Project administration: Milou S. C. Sep, R. Angela Sarabdjitsingh.

Supervision: Milou S. C. Sep, R. Angela Sarabdjitsingh, Marian Joëls.

Visualization: Milou S. C. Sep.

Writing – original draft: Milou S. C. Sep, Marijn Vellinga, Marian Joëls.

Writing – review & editing: R. Angela Sarabdjitsingh.

References

1. Rudy JW. Context representations, context functions, and the parahippocampal-hippocampal system. *Learn Mem.* 2009; 16: 573–585. <https://doi.org/10.1101/lm.1494409> PMID: 19794181
2. Maren S, Phan KL, Liberzon I. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nat Publ Gr.* 2013; 14: 417–428. <https://doi.org/10.1038/nrn3492> PMID: 23635870
3. Godden DR, Baddeley AD. Context-dependent memory in two natural environments: on land and underwater. *Br J Psychol.* 1975; 66: 325–331. <https://doi.org/10.1111/j.2044-8295.1975.tb01468.x>
4. Smith SM, Vela E. Environmental context-dependent memory: A review and meta-analysis. *Psychon Bull Rev.* 2001; 8: 203–220. <https://doi.org/10.3758/bf03196157> PMID: 11495110
5. De Quervain D, Schwabe L, Roozendaal B. Stress, glucocorticoids and memory: implications for treating fear-related disorders. *Nat Publ Gr.* 2017; 18: 7–19. <https://doi.org/10.1038/nrn.2016.155> PMID: 27881856
6. Lissek S, Rabin S, Heller RE, Lukenbaugh D, Geraci M, Pine DS, et al. Overgeneralization of Conditioned Fear as a Pathogenic Marker of Panic Disorder. *Am J Psychiatry.* 2010; 167: 47–55. <https://doi.org/10.1176/appi.ajp.2009.09030410> PMID: 19917595
7. El M, Roy HA-C, Kessels PC. Context Memory in Alzheimer's Disease. 2013 [cited 27 Jan 2021]. <https://doi.org/10.1159/000354187> PMID: 24403906
8. Urcelay GP, Miller RR. The functions of contexts in associative learning. *Behavioural Processes.* Elsevier; 2014. pp. 2–12. <https://doi.org/10.1016/j.beproc.2014.02.008> PMID: 24614400
9. Sep MSC, van Ast VA, Gorter R, Joëls M, Geuze E. Time-dependent effects of psychosocial stress on the contextualization of neutral memories. *Psychoneuroendocrinology.* 2019; 108: 140–149. <https://doi.org/10.1016/j.psyneuen.2019.06.021> PMID: 31280058
10. Zhang W, van Ast VA, Klumpers F, Roelofs K, Hermans EJ. Memory Contextualization: The Role of Prefrontal Cortex in Functional Integration across Item and Context Representational Regions. *J Cogn Neurosci.* 2018; 30: 579–593. https://doi.org/10.1162/jocn_a_01218 PMID: 29244638
11. Clewett D, DuBrow S, Davachi L. Transcending time in the brain: How event memories are constructed from experience. *Hippocampus.* 2019; 29: 162–183. <https://doi.org/10.1002/hipo.23074> PMID: 30734391
12. Ezzyat Y, Davachi L. Similarity Breeds Proximity: Pattern Similarity within and across Contexts Is Related to Later Mnemonic Judgments of Temporal Proximity. *Neuron.* 2014; 81: 1179–1189. <https://doi.org/10.1016/j.neuron.2014.01.042> PMID: 24607235
13. Clewett D, Gasser C, Davachi L. Pupil-linked arousal signals track the temporal organization of events in memory. *Nat Commun.* 2020; 11: 4007. <https://doi.org/10.1038/s41467-020-17851-9> PMID: 32782282
14. Libby LA, Reagh ZM, Bouffard NR, Ragland JD, Ranganath C. The Hippocampus Generalizes across Memories that Share Item and Context Information. *J Cogn Neurosci.* 2019; 31: 24–35. https://doi.org/10.1162/jocn_a_01345 PMID: 30240315
15. Staudigl T, Hanslmayr S. Theta Oscillations at Encoding Mediate the Context-Dependent Nature of Human Episodic Memory. *Curr Biol.* 2013; 23: 1101–1106. <https://doi.org/10.1016/j.cub.2013.04.074> PMID: 23746635
16. van Ast VA, Cornelisse S, Meeter M, Kindt M. Cortisol mediates the effects of stress on the contextual dependency of memories. *Psychoneuroendocrinology.* 2014; 41: 97–110. <https://doi.org/10.1016/j.psyneuen.2013.12.007> PMID: 24495611
17. Yonelinas AP, Ranganath C, Ekstrom AD, Wiltgen BJ. A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nat Rev Neurosci.* 2019; 20: 364–375. <https://doi.org/10.1038/s41583-019-0150-4> PMID: 30872808
18. Dix SL, Aggleton JP. Extending the spontaneous preference test of recognition: Evidence of object-location and object-context recognition. *Behav Brain Res.* 1999. [https://doi.org/10.1016/s0166-4328\(98\)00079-5](https://doi.org/10.1016/s0166-4328(98)00079-5) PMID: 10512585
19. Good MA, Barnes P, Staal V, McGregor A, Honey RC. Context- but not familiarity-dependent forms of object recognition are impaired following excitotoxic hippocampal lesions in rats. *Behav Neurosci.* 2007; 121: 218–23. <https://doi.org/10.1037/0735-7044.121.1.218> PMID: 17324066

20. Davachi L. Item, context and relational episodic encoding in humans. *Curr Opin Neurobiol.* 2006; 16: 693–700. <https://doi.org/10.1016/j.conb.2006.10.012> PMID: 17097284
21. La Spina M, Sansevero G, Biasutto L, Zoratti M, Peruzzo R, Berardi N, et al. Pterostilbene Improves Cognitive Performance in Aged Rats: An in Vivo Study. *Cell Physiol Biochem.* 2019; 52: 232–239. <https://doi.org/10.33594/000000017> PMID: 30816671
22. De Rosa R, Garcia AA, Braschi C, Capsoni S, Maffei L, Berardi N, et al. Intranasal administration of nerve growth factor (NGF) rescues recognition memory deficits in AD11 anti-NGF transgenic mice. *Proc Natl Acad Sci U S A.* 2005; 102: 3811–6. <https://doi.org/10.1073/pnas.0500195102> PMID: 15728733
23. Rodberg EM, den Hartog CR, Anderson RI, Becker HC, Moorman DE, Vazey EM. Stress Facilitates the Development of Cognitive Dysfunction After Chronic Ethanol Exposure. *Alcohol Clin Exp Res.* 2017; 41: 1574–1583. <https://doi.org/10.1111/acer.13444> PMID: 28753742
24. Pillai AG, Arp M, Velzing E, Lesuis SL, Schmidt M V, Holsboer F, et al. Early life stress determines the effects of glucocorticoids and stress on hippocampal function: Electrophysiological and behavioral evidence respectively. *Neuropharmacology.* 2018; 133: 307–318. <https://doi.org/10.1016/j.neuropharm.2018.02.001> PMID: 29412144
25. Ennaceur A, Delacour J. A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behav Brain Res.* 1988; 31: 47–59. [https://doi.org/10.1016/0166-4328\(88\)90157-x](https://doi.org/10.1016/0166-4328(88)90157-x) PMID: 3228475
26. Ennaceur A. One-trial object recognition in rats and mice: methodological and theoretical issues. *Behav Brain Res.* 2010; 215: 244–54. <https://doi.org/10.1016/j.bbr.2009.12.036> PMID: 20060020
27. de Vries RBM, Hooijmans CR, Langendam MW, van Luijk J, Leenaars M, Ritskes-Hoitinga M, et al. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclin Med.* 2015; 2: e00007. <https://doi.org/10.1002/ebm2.7>
28. Leenaars M, Hooijmans CR, van Veggel N, ter Riet G, Leeflang M, Hooft L, et al. A step-by-step guide to systematically identify all relevant animal studies. *Lab Anim.* 2012; 46: 24–31. <https://doi.org/10.1258/la.2011.011087> PMID: 22037056
29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* 2009; 6: e1000097. <https://doi.org/10.1371/journal.pmed.1000097> PMID: 19621072
30. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. Boutron I, editor. *PLOS Biol.* 2020; 18: e3000410. <https://doi.org/10.1371/journal.pbio.3000410> PMID: 32663219
31. Vellinga M, Sep MSC, Joëls M, Geuze E. Measuring context-dependent memory in rodents: a systematic review and meta-analysis of important variables in the object in context-task (CRD42020191340). PROSPERO. 2020. Available: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020191340
32. Huwaldt JA. Plot Digitizer. Source Forge. 2015. Available: <http://plotdigitizer.sourceforge.net/>
33. Antunes M, Biala G. The novel object recognition memory: neurobiology, test procedure, and its modifications. *Cogn Process.* 2012; 13: 93–110. <https://doi.org/10.1007/s10339-011-0430-z> PMID: 22160349
34. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYR-CLE's risk of bias tool for animal studies. *BMC Med Res Methodol.* 2014; 14: 43. <https://doi.org/10.1186/1471-2288-14-43> PMID: 24667063
35. Bonapersona V, Kentrop J, Van Lissa CJ, van der Veen R, Joëls M, Sarabdjitsingh RA. The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies. *Neurosci Biobehav Rev.* 2019; 102: 299–307. <https://doi.org/10.1016/j.neubiorev.2019.04.021> PMID: 31047892
36. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2020. Available: <https://www.r-project.org/>
37. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.2. 2020.
38. Wolen A, Hartgerink C, Hafen R, Richards B, Soderberg C, York T. osfr: An R Interface to the Open Science Framework. *J Open Source Softw.* 2020; 5: 2071. <https://doi.org/10.21105/joss.02071>
39. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw.* 2010; 36: 1–48. <https://doi.org/10.18637/jss.v036.i03>
40. van Lissa CJ. metaforest: Exploring Heterogeneity in Meta-Analysis using Random Forests. R package version 0.1.3. 2020.
41. Kuhn M. caret: Classification and Regression Training. R package version 6.0–86. 2020. Available: <https://cran.r-project.org/package=caret>

42. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.
43. Borenstein M, Hedges L V, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. 1st ed. West Sussex: John Wiley & Sons, Ltd; 2009.
44. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *British Medical Journal*. 2003. <https://doi.org/10.1136/bmj.327.7414.557> PMID: 12958120
45. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull*. 1979. <https://doi.org/10.1037/0033-2909.86.3.638>
46. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000. <https://doi.org/10.1111/j.0006-341X.2000.00455.x> PMID: 10877304
47. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997; 315: 629–34. <https://doi.org/10.1136/bmj.315.7109.629> PMID: 9310563
48. Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*. 1994. <https://doi.org/10.2307/2533446>
49. Viechtbauer W, Cheung MW-L. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods*. 2010. <https://doi.org/10.1002/jrsm.11> PMID: 26061377
50. Apley DW, Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. *J R Stat Soc Ser B Stat Methodol*. 2020; 82: 1059–1086. <https://doi.org/10.1111/rssb.12377>
51. Higgins J, Li T, Deeks J. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions version 61*. 2020. Available: www.training.cochrane.org/handbook.
52. Cohen J. *Statistical power analysis for the behavioural sciences*. New York: Academic Press; 1969.
53. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007; 39: 175–191. <https://doi.org/10.3758/bf03193146> PMID: 17695343
54. Kesner RP. An analysis of dentate gyrus function (an update). *Behav Brain Res*. 2018; 354: 84–91. <https://doi.org/10.1016/j.bbr.2017.07.033> PMID: 28756212
55. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLoS Biol*. 2010; 8: e1000412. <https://doi.org/10.1371/journal.pbio.1000412> PMID: 20613859
56. Antonic A, Sena ES, Lees JS, Wills TE, Skeers P, Batchelor PE, et al. Stem Cell Transplantation in Traumatic Spinal Cord Injury: A Systematic Review and Meta-Analysis of Animal Studies. Altman DG, editor. *PLoS Biol*. 2013; 11: e1001738. <https://doi.org/10.1371/journal.pbio.1001738> PMID: 24358022
57. Verbitsky A, Dopfel D, Zhang N. Rodent models of post-traumatic stress disorder: behavioral assessment. *Transl Psychiatry*. 2020; 10: 132. <https://doi.org/10.1038/s41398-020-0806-x> PMID: 32376819
58. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014; 383: 166–175. [https://doi.org/10.1016/S0140-6736\(13\)62227-8](https://doi.org/10.1016/S0140-6736(13)62227-8) PMID: 24411645
59. Kilkenny C, Parsons N, Kadyszewski E, Festing MFW, Cuthill IC, Fry D, et al. Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals. McLeod M, editor. *PLoS One*. 2009; 4: e7824. <https://doi.org/10.1371/journal.pone.0007824> PMID: 19956596
60. Watzlawick R, Antonic A, Sena ES, Kopp MA, Rind J, Dirnagl U, et al. Outcome heterogeneity and bias in acute experimental spinal cord injury: A meta-analysis. *Neurology*. 2019; 93: e40–e51. <https://doi.org/10.1212/WNL.00000000000007718> PMID: 31175207
61. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The Extent and Consequences of P-Hacking in Science. *PLOS Biol*. 2015; 13: e1002106. <https://doi.org/10.1371/journal.pbio.1002106> PMID: 25768323
62. Van Lissa C. MetaForest: Exploring heterogeneity in meta-analysis using random forests. 2017. <https://doi.org/10.31234/osf.io/myg6s>
63. Mumby DG. Hippocampal Damage and Exploratory Preferences in Rats: Memory for Objects, Places, and Contexts. *Learn Mem*. 2002; 9: 49–57. <https://doi.org/10.1101/lm.41302> PMID: 11992015
64. Bonapersona V, Hoijtink H, Consortium R, Sarabdjitsingh RA, Joëls M. Increasing the statistical power of animal experiments with historical control data. *Nat Neurosci*. 2021. <https://doi.org/10.1038/s41593-020-00792-3> PMID: 33603229
65. Blaser R, Heyser C, Parker MO, Van Der Zee EA, Hyun Kim J. Spontaneous object recognition: a promising approach to the comparative study of memory. *Front Behav Neurosci*. 2015; 9: 183. <https://doi.org/10.3389/fnbeh.2015.00183> PMID: 26217207

66. Binder S, Dere E, Zlomuzica A. A critical appraisal of the what-where-when episodic-like memory test in rodents: Achievements, caveats and future directions. *Prog Neurobiol*. 2015; 130: 71–85. <https://doi.org/10.1016/j.pneurobio.2015.04.002> PMID: 25930683
67. Gökçek-Saraç Ç, Wesierska M, Jakubowska-Doğru E. Comparison of spatial learning in the partially baited radial-arm maze task between commonly used rat strains: Wistar, Sprague-Dawley, Long-Evans, and outcrossed Wistar/Sprague-Dawley. *Learn Behav*. 2015; 43: 83–94. <https://doi.org/10.3758/s13420-014-0163-9> PMID: 25537841
68. Sep MSC, Joëls M, Geuze E. Individual differences in the encoding of contextual details following acute stress: An explorative study. *Eur J Neurosci*. 2020; ejn.15067. <https://doi.org/10.1111/ejn.15067> PMID: 33249674
69. Imuta K, Scarf D, Carson S, Hayne H. Children's learning and memory of an interactive science lesson: Does the context matter? *Dev Psychol*. 2018; 54: 1029–1037. <https://doi.org/10.1037/dev0000487> PMID: 29251971
70. Powell PS, Strunk J, James T, Polyn SM, Duarte A. Decoding selective attention to context memory: An aging study. *Neuroimage*. 2018; 181: 95–107. <https://doi.org/10.1016/j.neuroimage.2018.06.085> PMID: 29991445
71. Strunk J, James T, Arndt J, Duarte A. Age-related changes in neural oscillations supporting context memory retrieval. *Cortex*. 2017; 91: 40–55. <https://doi.org/10.1016/j.cortex.2017.01.020> PMID: 28237686
72. Ankudowich E, Pasvanis S, Rajah MN. Age-related differences in prefrontal-hippocampal connectivity are associated with reduced spatial context memory. *Psychol Aging*. 2019; 34: 251–261. <https://doi.org/10.1037/pag0000310> PMID: 30407034
73. Kirschen GW, Ge S. Young at heart: Insights into hippocampal neurogenesis in the aged brain. *Behav Brain Res*. 2019; 369: 111934. <https://doi.org/10.1016/j.bbr.2019.111934> PMID: 31054278
74. Alam MJ, Kitamura T, Saitoh Y, Ohkawa N, Kondo T, Inokuchi K. Adult neurogenesis conserves hippocampal memory capacity. *J Neurosci*. 2018; 38: 6854–6863. <https://doi.org/10.1523/JNEUROSCI.2976-17.2018> PMID: 29986876
75. Andreano JM, Cahill L. Sex influences on the neurobiology of learning and memory. *Learn Mem*. 2009; 16: 248–266. <https://doi.org/10.1101/lm.918309> PMID: 19318467
76. Keeley RJ, Bye C, Trow J, McDonald RJ. Strain and sex differences in brain and behaviour of adult rats: Learning and memory, anxiety and volumetric estimates. *Behav Brain Res*. 2015; 288: 118–131. <https://doi.org/10.1016/j.bbr.2014.10.039> PMID: 25446747
77. Yagi S, Galea LAM. Sex differences in hippocampal cognition and neurogenesis. *Neuropsychopharmacology*. 2019; 44: 200–213. <https://doi.org/10.1038/s41386-018-0208-4> PMID: 30214058
78. Diamond DM, Campbell AM, Park CR, Halonen J, Zoladz PR. The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plast*. 2007; 2007: 1–33. <https://doi.org/10.1155/2007/60803> PMID: 17641736
79. Quaedflieg CWEM, Schwabe L. Memory dynamics under stress. *Memory*. 2018; 26: 364–376. <https://doi.org/10.1080/09658211.2017.1338299> PMID: 28625108
80. Rao RP, Anilkumar S, McEwen BS, Chattarji S. Glucocorticoids Protect Against the Delayed Behavioral and Cellular Effects of Acute Stress on the Amygdala. *BPS*. 2012; 72: 466–475. <https://doi.org/10.1016/j.biopsych.2012.04.008> PMID: 22572034