Draft version 2024-01-10 submitted for peer-review

Pupil Old/New Effect as an Objective Measure of Memory: A Meta-Analysis of 17 Eye-Tracking Experiments

Alexandra Lapteva & Corinna S. Martarelli

Faculty of Psychology, UniDistance Suisse, Switzerland

Correspondence: alexandra.lapteva@unidistance.ch

Pupil Old/New Effect as an Objective Measure of Memory: A Meta-Analysis of 17 Eye-Tracking Experiments

Abstract

Recognition memory, the ability to recognize previously encountered information, correlates with pupil diameter changes during the recognition period. This physiological response, known as the pupil old/new effect, reflects the variation in pupil dilation when encountering previously studied (old) stimuli compared to new objects. To develop a more precise understanding of the pupil/old-new effect, we conducted a meta-analysis of 17 eye-tracking experiments (across 12 articles spanning from 2008 to 2023) involving 560 healthy adults with a mean age of 22.31 years. The main analysis revealed a significant and large pooled pupil old/new effect (Cohen's d = 0.85, 95% CI [0.48, 1.21]). Further sensitivity analysis of the moderators showed that the memory type (incidental/intentional), time window of the pupil response (which accounted for 72.82% of the heterogeneity of the experiments included in the meta-analysis), sampling rate, and pupil old/new effect measure each had a significant impact on the meta-analytic results. Additionally, an analysis of bias showed a rather low risk of bias in the selected articles. In general, the analyses revealed a robust pupil old/new effect across all selected articles. This finding underscores its potential utility as a marker of memory recognition across stimulus types, including visual/auditory and neutral/emotional stimuli.

Keywords: pupil old/new effect, pupil dilation, recognition, meta-analysis

Pupil Old/New Effect as an Objective Measure of Memory: A Meta-Analysis of 17 Eye-Tracking Experiments

Introduction

The experience of recognizing a familiar face in a crowd of strangers or a melody that we have not heard in a while is a common occurrence. While our brain navigates between old and new stimuli every day and dissociates them unobtrusively, how this process of recognition is represented in our physiological reactions (e.g., in terms of pupil diameter changes) is in need of further exploration. Studying recognition provides a window into the interplay between memory and perception, as well as the cognitive mechanisms underlying the process of dissociating old and new stimuli. The meta-analysis presented in this article was thus designed to investigate the study of the pupil old/new effect as a measure of recognition memory and to synthesize and summarize existing research on the topic. In the following paragraphs, we first discuss the historical background of the research that discovered the pupil old/new effect, uncover the theories that have been proposed to explain this phenomenon, and, finally, introduce the meta-analysis we conducted to explore how the pupil old/new effect can be understood as an objective measure of memory.

Pupillometry and Memory Recognition: Historical Background

Recognition memory refers to the ability to recognize previously encountered information (Rugg & Curran, 2007). Within the framework of recognition memory, interesting insights into our mental activity have emerged through the investigation of pupillometry, which refers to the study of pupil diameter changes in response to cognitive processing (Sirois & Brisson, 2014). Specifically, such changes have been associated with certain components of recognition memory research, such as the identification and judgment of episodic content (Brown & Xiang, 1998), mental effort (Alnaes et al., 2014; Kahneman & Beatty, 1966), recollection and familiarity processes (Kafkas & Montaldi, 2011; Naber et al., 2013; Papesh et al., 2012; Võ et al., 2008), and attention (Binda et al., 2014; Naber et al., 2013; Wang & Munoz, 2015).

Pioneering investigations combining pupil size and memory have been conducted with regards to the topic of mental load; in 1966, Kahneman and Beatty presumed that pupil diameter was a measure of the active processing of material during a short-term memory task (Kahneman & Beatty, 1966, replicated in Beatty, 1982, and Kucewicz et al., 2018). The experiments conducted by Gardner, Mo, and Borrego (1974) and Gardner, Mo, and Krinsky (1974) further advanced research on the pupil size changes induced by mental load by investigating the connection between recognition memory and the pupillary orienting reflex (also referred to as the "pupillary light reflex"). Not only were these researchers among the first to demonstrate different pupillary reactions that occur when viewing old and novel stimuli, but they also laid out the groundwork for exploring the reasons behind this phenomenon by questioning whether task difficulty and task requirements play a major role in such pupil reactions.

A shift toward investigating old-new differences emerged with the rise of event-related potential (ERP) research (e.g., Duarte et al., 2006; Curran & Cleary, 2003; Rugg & Curran, 2007). For instance, in a study design consisting of both ERP and behavioral measures, Duarte et al. (2006) were able to detect the difference between old (remembered from the preceding test session), familiar (known), and new items and to match the information with the ERPs as a double measure of memory. After the popularization of ERP research on the old/new effect, Võ et al. (2008) successfully adapted this methodology to pupillometry, circling back to the pioneering research of the 1960s. They found a distinct old/new effect in the pupil size for old and new stimuli at recognition, subsequently labeling this the "pupil old/new effect," which has been replicated in other experiments (e.g., Brocher & Graf, 2016; Otero et al., 2011; Papesh et al., 2012). Thus, the pupil old/new effect is described in the literature as the phenomenon of a larger pupil size being observed in connection with old stimuli and a smaller pupil size for new stimuli at memory recognition.

Explanatory Theories for the Pupil Old/New Effect

Several theoretical frameworks have been advanced to explain the reasons behind the larger pupillary dilation to old versus new stimuli. Initially, and consistent with the research that emerged earlier in the 20th century (e.g., Kahneman & Beatty, 1966), Võ et al. (2008), the pioneers of this field of research, argued that the pupil old/new effect is modulated by the cognitive load that goes into encoding an item. In their view, smaller pupil dilation during recognition indicates less cognitive load being employed during encoding, while larger pupil dilation during recognition reflects a stronger cognitive load employed during encoding. Furthermore, they found different pupil sizes for the emotional and neutral recognition of old words. Specifically, emotional stimuli elicited smaller pupil dilations than neutral stimuli. Accordingly, Võ et al. argued that less cognitive effort goes into encoding emotional stimuli, as emotional stimuli possess some processing advantage compared to neutral stimuli.

As a challenge to the cognitive load hypothesis, researchers such as Otero et al. (2011) have suggested that the depth of encoding is a superior explanation for the pupil old/new effect. They argue that the pupil old/new effect is modulated by the strength of the memory upon which the recognition judgment is based. Contrary to Võ et al.'s (2008) emphasis on cognitive load and effort, Otero et al. (2011) have found a more pronounced pupil old/new effect for items encoded under deep versus shallow conditions, and they have also found that the effect presents itself for false recognitions (i.e., new items erroneously identified as old). Otero et al. (2011) have thus suggested that the variety of pupillary reactions to different stimuli and encoding instructions stems from the combined signals of both recollection and familiarity processes that influence pupil dilation and constriction. Building on this notion, Heaver and Hutton (2011) have further suggested that confidence in recognition judgments reflects the strength of a memory trace, thus influencing the pupil old/new effect. Specifically, they found that the pupil old/new effect was present even when participants were instructed to deliberately misidentify old items as new, thereby feigning amnesia.

In addition, some researchers have refined the previous ideas of Võ et al. and considered the process of encoding as a means of explaining the subsequent pupil old/new effect. Papesh et al. (2012) found that greater cognitive effort that goes into encoding (which is reflected by bigger pupil dilation at encoding) is associated with a more pronounced pupil dilation during recognition. As previous researchers (Heaver & Hutton, 2011) hypothesized, Papesh et al. found that the pupil old/new effect varies across different confidence levels, therefore suggesting that the pupil size reflects the strength and contents of memory. Later research, such as the work of Brocher and Graf (2016, 2017), has concluded that the pupil old/new effect is connected to a participant's focus of attention at encoding. In their view, it is important to look at the way the experimental task is designed, specifically the task demands and the participant's attentional engagement in the task itself, as these factors might shape the magnitude of the pupil old/new effect. Primarily, Brocher and Graf's research results associate the pupil old/new effect with the controlled and conscious use of memory, rather than an incidental function.

Present Meta-Analysis

Despite the existence of multiple theories explaining the effect and a substantial database of experiments on the subject, a comprehensive meta-analysis focusing on the pupil old/new effect of memory recognition remains absent in the published literature. Our goal was thus to address this gap by providing a quantitative summary of previous eye-tracking experiments centered on recognition memory.

There is some variability in the research devoted to the pupil old/new effect. For example, empirical findings vary regarding the nature of the stimuli used in experimental designs. For instance, the emotional valence of stimuli presented during experiments has been reported to modulate pupil size. As previously mentioned, Võ et al. (2008) found that the pupil old/new effect was the most prominent for neutral stimuli compared to the diminished effect for emotional stimuli (see also Brocher & Graf, 2016, Experiment 2). However, Bradley and Lang (2015) found more pronounced pupil dilations for old emotional stimuli than for old neutral stimuli. Another source of potential inconsistency in the research into the pupil old/new effect is the experimental instructions researchers use for their experiments, and, in particular, whether participants know if a memory test follows the encoding of the to-be-remembered material (i.e., whether the encoding is intentional or incidental). This could lead to varying results due to the amount of cognitive effort exerted to remember stimuli, as focused attention during intentional encoding can elicit better memory performance (Tyler et al., 1979), while other confounding factors, such as mind-wandering during incidental encoding (e.g., elicited by thought associations), can yield worsened memory performance (Maillet & Schacter, 2016). Another source of variability in the research is the option to include all trials (both correct and incorrect; e.g., Herweg et al., 2018; Kafkas & Montaldi, 2015) or only correct trials (e.g., Bradley & Lang, 2015; Heaver & Hutton, 2011) in the data analysis.

By gathering data from multiple articles that included visual and/or auditory stimuli, different valences of stimuli (emotional/neutral), different experimental instructions for the encoding of stimuli (incidental/intentional), and estimations of the overall effect size of the phenomenon, our aim was to establish a robust set of evidence and a solid research database for the pupil old/new effect. While anticipating the presence of a pupil old/new effect, we expected inherent heterogeneity in the findings, due to variations in experimental design, data analysis approaches, and formulas used to measure the effect. Accordingly, we identified precise inclusion criteria for the articles that were considered in this meta-analysis, fitted forest and funnel plots to portray the meta-analytic results, performed overall risk of bias and publication bias assessments, evaluated the potential moderators of the effect, and, finally, conducted an outlier analysis to establish a more accurate effect size of the pupil old/new effect as an objective measure of memory.

Method

Eligibility: Inclusion and Exclusion Criteria

The focus of this meta-analysis was on articles that tested the pupil old/new effect and provided sufficient pupil data from the recognition memory tasks of the experiments. Therefore, we chose the following inclusion criteria for the identified articles: 1) they needed to empirically test the pupil old/new effect, specifically with the aim of examining the difference in the pupil size between old and new stimuli in a recognition memory task; 2) they had to include pupil measurements for the recognition phase of the experiment; 3) they had to contain sufficient statistical information to compute Cohen's *d* (means, standard deviations, *F* statistics, *T* statistics, and sample sizes); 4) they had to be published in peer-reviewed journals; and 5) they had to be written in English. Articles that did not meet these criteria were not included in the meta-analysis. For example, articles containing insufficient statistics (e.g., partial eta-squared as an effect size or model comparisons) were excluded from the analysis because we were unable to calculate Cohen's *d*.

Literature Search

We used the PRISMA statement guidelines for carrying out this meta-analysis (Page et al., 2021). The systematic literature search was completed in the spring of 2023 by two separate researchers and was limited to articles published before May 2023. We carried out the literature search in Scopus, Google Scholar, PubMed, and several other web databases, including ClinicalTrials.gov, WHO's International Clinical Trials Registry Platform (ICTRP), OpenDOAR (Directory of Open-Access Repositories), and WorldCat. The following keywords were used for the search: [pupil old/new effect], [pupil old new effect], [retrieval memory and pupillometry], [recollection and pupil size], and [pupillometry and familiarity]. We looked for articles in peer-reviewed journals and in the reference lists of retrieved articles.

Data Extraction

Considering the diversity of experimental designs used in the literature, we extracted the data based on characteristics that we commonly identified across articles that featured the pupil old/new effect. These characteristics included 1) the number of participants; 2) memory type (incidental/intentional); 3) type of stimuli (pictures/words); 4) stimuli valence (emotional/neutral); 5) trial type, based on which the pupil old/new effect was calculated (all/correct); 6) eye-tracker hardware; 7) sampling rate of the eye-tracking; 8) preprocessing of the pupil data; 9) baseline of the pupil size used to calculate the pupil old/new effect; 10) time window of the pupil measures (i.e., baseline and trial locked period for measuring the pupil response); and 11) exact pupil old/new effect calculation (see Table 1 for the summary). These characteristics will be referenced later to account for the variability in the effect sizes across the articles.

Meta-Analytic Procedure

Effect Size Calculation

We chose Cohen's *d* as the effect size measure (Cohen, 1988). It is important to note that the vast majority of meta-analyses typically focus on analyzing intervention or treatment effects, thus involving between-group comparisons (Barendregt et al., 2013). Accordingly, the majority of such studies in the existing literature and the established guidelines for conducting meta-analyses and effect size calculations are based on between-group measures. In this case, we sought to conduct a meta-analysis of research on within-group differences in pupil size between new and old stimuli during recognition memory. As all of the articles included in the meta-analysis involved repeated measures designs, and the available resources that are specifically tailored for conducting meta-analyses for within-groups are scarce, we relied on the guidelines for between-group measures. This was advantageous in our case, as between-group effect size measures tend to yield a more conservative estimation than within-group effect size measures. This is because the existence of a correlation between measures in repeated designs can reduce the variability and sometimes lead to inflated effect sizes (e.g., Cohen's d_z or eta squared/omega squared; Dunlap et al., 1996; Olejnik & Algina, 2003). According to Loftus and Masson (1994), inter-subject variance in a within-subject group design can be ignored because of its statistical irrelevance; thus, choosing a more conservative option of disregarding the correlation within the group was considered more beneficial for the results of the present meta-analysis.

All statistical analyses were completed in R version 4.3.1 (R Core Team, 2020) with the packages meta (Balduzzi et al., 2019), metafor (Viechtbauer, 2010) and weightr (Coburn & Vevea, 2015). All scripts and data analyses were shared via the Open Science Foundation OSF (https://osf.io/5pgcj/?view_only=51e59437d86e4cc8b1bff4e582f25b8c). We computed Cohen's *d* for each study to compare the magnitude of the difference between pupil size for old and new stimuli during the recognition memory task. Cohen's *d* was calculated with statistics from the chosen articles with the esc package for R designed for between-measures effect size calculation (Lüdecke, 2018). Effect sizes were calculated separately for each experimental result in articles containing more than one experiment. Cohen's *d* was calculated with the standard formula $d = (ME - MC)/SD$ and interpreted according to Cohen's guidelines in terms of the following sizes: small (*d* = 0.2), medium (*d* = 0.5), and large (*d* = 0.8; Cohen, 1988).

Risk of Bias Assessment

To assess the methodological quality of the articles included in the meta-analysis, we performed a risk of bias assessment on the effect sizes with the Cochrane Risk of Bias 2 tool (RoB, Sterne et al., 2019). The assessment was conducted by two independent researchers. The RoB tool includes five domains against which articles are examined for publication bias: D1) the randomization process involved, D2) deviations from the intended interventions, D3) missing outcome data, D4) measurement of the outcome, and D5) selection of the reported results. The algorithm evaluates assessor's answers to the questions included in each domain and outputs the "low", "some concerns" or "high" risk of bias judgements for each experiment in the meta-analysis. The overall risk of bias is then displayed as the worst risk of bias in any evaluated domain (Sterne et al., 2019).

We adjusted the tool by adapting some of the questions to fit the within-measures design of the articles included in the meta-analysis. D1, which assesses the randomization process, was deemed unnecessary, as the focus of our analysis was on the differential pupil responses within experimental manipulations rather than between-group interventions. Regarding D2, which pertains to bias due to deviations from intended interventions, none of the articles included in the meta-analysis reported deviations from the intended protocol, and D2 was deemed irrelevant to our research as well. Articles were then assessed in terms of the remaining three domains: D3, D4, and D5. The risk of bias in D3 was evaluated with respect to how the researchers described missing outcome data (e.g., if the missing data were extrapolated, or participants with missing data were excluded). The risk of bias according to D4 was evaluated by how the researchers measured the outcome, which in our case referred to how the pupil old/new effect was measured. Finally, the risk of bias according to D5 was assessed in terms of whether the researchers reported the results according to the prespecified analysis.

Publication Bias Assessment

Publication bias can influence the results of a meta-analysis; it originates from the tendency of researchers to publish articles depending on the obtained results, as articles with statistically significant results are more likely to be published (Fanelli, 2010; Rosenthal, 1979). We used several methods to assess publication bias in the selected articles. We fitted a funnel plot (Egger et al., 1997), which represents effect sizes against the sample sizes in each experiment, to assess the validity of the meta-analysis and detect potential asymmetry. If all results are symmetrically distributed around the average effect size, the funnel plot suggests that there is no bias. If the plot is not symmetrical, it might indicate publication bias or other factors influencing the results. Asymmetry can be caused by various factors, such as small-study effects (Sterne et al., 2000), which can be explained by experiments with small sample sizes showing larger effects compared to larger sample sizes. Small sample sizes can lead to increased variability and thus exaggerated effect sizes. To assess funnel plot asymmetry caused by small-study effects, we ran Egger's test of the intercept.

To address publication bias, we further fitted a selection model test, which is a class of techniques proposed by Hedges (1984), that help adjust the effect size and calculate it without publication bias. We chose a three-parameter model, which is a selection model with only three estimated parameters: the effect size parameter (i.e., the average effect size across experiments), the heterogeneity parameter (i.e., the variability between the effect sizes of individual experiments), and the weight parameter (i.e., the weight given to each experiment in calculating the overall effect size; McShane et al., 2016). The model assesses and adjusts publication bias and extracts the true effect size.

Outlier Analysis

To provide more robust evidence of the pupil old/new effect size, we performed an outlier analysis to detect experiments that had a disproportionate influence on the effect size. Outliers were detected if their individual 95% confidence intervals (CI) were outside the 95% CI of the pooled effect. The effect size with outliers removed was calculated and compared with the initial effect size.

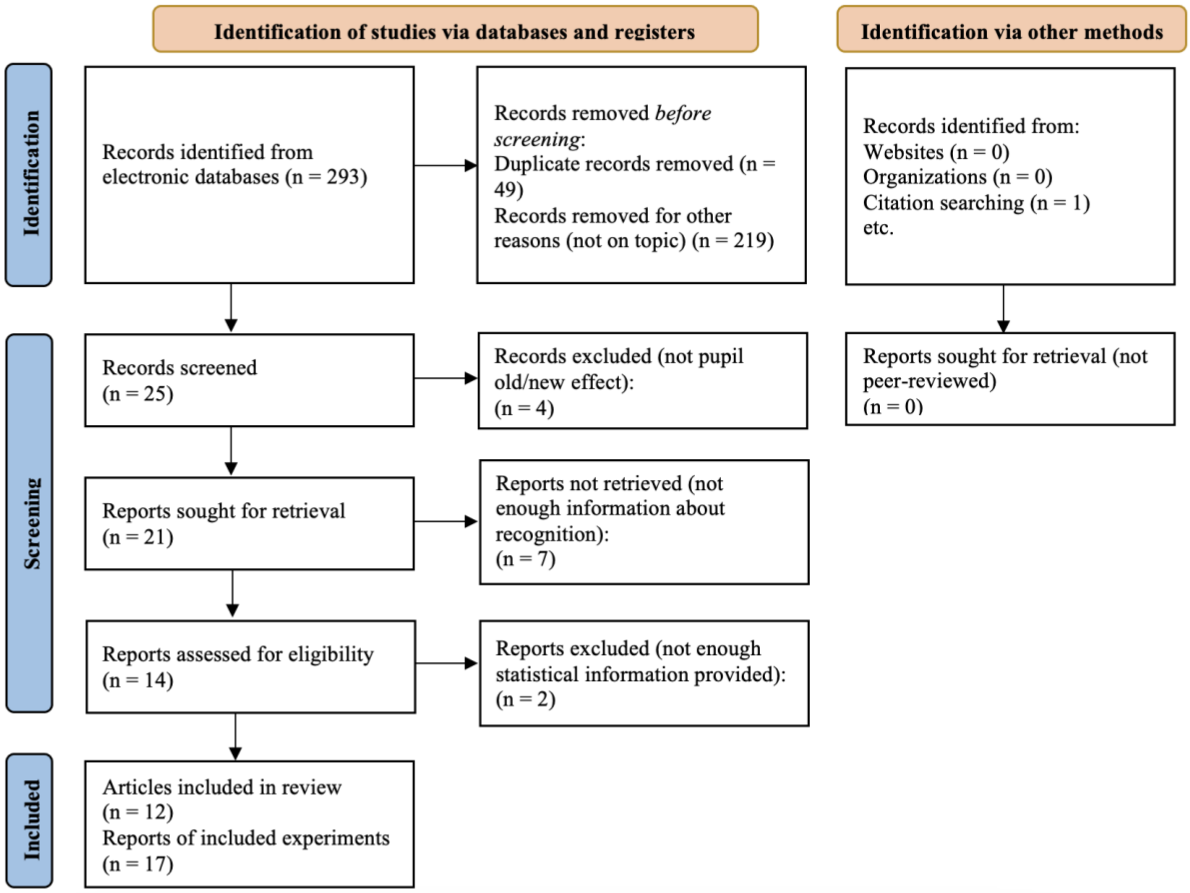
Results

Article Selection

The literature search (see Figure 1) is summarized in terms of the PRISMA flow chart (Page et al., 2021). The initial keyword search resulted in 293 articles. After deleting the duplicates (*n* = 49) and screening the abstracts for articles that were not on the topic (*n* = 219), we were left with 25 articles that were assessed for eligibility criteria. Finally, 12 articles with 17 experiments were included in the meta-analysis.

Figure 1

PRISMA Flow Chart of the Literature Search



Article Characteristics

Data from a total of 17 experiments from 12 articles (participants n = 560) were selected for further meta-analytic calculations. These experiments were classified according to the 11 characteristics mentioned above (see "Data Extraction" and Table 1). All the experiments included in the analysis featured a general healthy adult population. To increase the number of articles that fit in our meta-analysis and potentially strengthen the validity of our findings, we included one article with cueing conditions (Mill et al., 2016), one article that featured shallow and deep encoding designs (Otero et al., 2011), and one article that investigated the pupil old/new effect in high confidence memory performance ratings (Heaver & Hutton, 2011). Incorporating experiments with different experimental designs allows for enhanced validity, as a consistent effect size, regardless of the experimental design, further proves the reliability of the results.

All 17 experiments featured within-group designs and were conducted with eye-tracking devices. Among these experiments, 15 out of 17 (88%) were conducted with university students; in contrast, the studies by Herweg et al. (2018) and Oliveira et al. (2021) involved the general population (mean age across all of the experiments = 22.31 years). In terms of memory encoding, seven out of 17 experiments (41%) featured incidental memory encoding, seven out of 17 (41%) focused on intentional memory encoding, and three out of 17 (18%) did not indicate the memory type. Regarding stimuli valence, the experiments featured emotional (positive/negative) and neutral stimuli. Only two out of 17 experiments (12%) investigated the pupil old/new effect for both emotional and neutral stimuli, while the remaining focused on neutral stimuli (88%). We also investigated the different types of stimuli used across experiments: five out of 17 experiments (29%) featured visual stimuli (pictures on a screen), nine (53%) featured words on a screen, two (12%) used auditory stimuli (nouns presented to participants via headphones), and one (6%) showed participants both words and pictures on a screen. Further, six out of 17 experiments (35%) featured information based on correct trials (hits and correct rejections), and 11 out of 17 experiments (65%) featured all trials (hits, correct rejections, false alarms, and misses).

Table 1

Articles Examining the Pupil Old/New Effect Included in the Meta-Analysis

	Authors	N of participants	Memory	Stimuli	Stimuli valence	Trials	Eye-tracker hardware	Sampling rate	Preprocessing of the pupil data	Baseline of the pupil size	Time window of the pupil response	Measure of old-new effect
1.	Bradley & Lang, 2015	65	Incidental	Pictures on the screen	Emotional (half erotica, half violence) and neutral	Correct	ASL model 504 eye tracker	60 HZ	Blinks removed, linear interpolation of missing data	0–1000 ms before trial	0–3000 ms (stimulus locked)	Baseline corrected peak dilation
2.	Campos-Arteaga et al., 2020	17	Incidental	Pictures with words on the screen	Neutral (both pictures and words)	All	Tobii Pro TX300	300 HZ	Blinks and signal loss removed, linear interpolation of missing data, down-sampling of data	0–1000 ms before trial	0–4000 ms (stimulus locked)	Baseline corrected maximal dilation
3.	Dobbins, 2021 Experiment 1	36	Intentional	Words on the screen	Neutral	Correct	EyeLink 1000	1000 HZ	Blinks and artifacts removed, linear interpolation of missing data, down-sampling of data	0–1000 ms during the trial	0–4000 ms (stimulus locked)	Baseline corrected mean dilation
4.	Dobbins, 2021 Experiment 2	43	Intentional	Words on the screen	Neutral	Correct	EyeLink 1000	1000 HZ	Blinks and artifacts removed, linear interpolation of missing data, down-sampling of data	0–1000 ms during the trial	0–4000 ms (stimulus locked)	Baseline corrected mean dilation

5.	Heaver & Hutton, 2011 1/10/2024 12:38:00 PM	26	Intentional	Words on the screen	Neutral	Correct	Eyelink II	500 HZ	Not stated	0–250 ms before trial	250–1750 ms	Pupil dilation ratio (PDR)
6.	Hellmer et al., 2018 Experiment 1	16	Incidental	Pictures on the screen	Neutral (everyday objects)	All	Tobii T120	60 HZ	Smoothing and interpolation of missing data	0–500 ms during the trial	1000–2000 ms	Baseline corrected mean dilation
7.	Herweg et al., 2018	24	Not stated	Pictures on the screen	Neutral (scene images)	All	EyeLink 1000	1000 HZ	Fixations, saccades and blinks identified, spline interpolation of data, down-sampling of data	0–240 ms before trial	0–2700 ms	Baseline corrected mean dilation
8.	Kafkas & Montaldi, 2015 Experiment 1	37	Incidental	Pictures on the screen	Neutral (natural and human-made objects)	All	ASL Model Eye-Trac 6000	60 HZ	Blinks and signal loss removed	0–1000 ms before trial	0–4000 ms (stimulus locked)	Baseline corrected peak pupil response
9.	Kafkas & Montaldi, 2015 Experiment 2	29	Incidental	Pictures on the screen	Neutral (natural and human-made objects)	All	ASL Model Eye-Trac 6000	60 HZ	Blinks and signal loss removed	0–1000 ms before trial	0–4000 ms (stimulus locked)	Baseline corrected peak pupil response
10.	Mill et al., 2016	34	Incidental	Words on the screen	Neutral	All	Eyelink 1000	1000 HZ	Blinks and signal loss removed, linear interpolation of missing data, down-sampling of data	0–200 ms before trial	0–2000 ms (stimulus locked)	Baseline corrected mean dilation
11.	Oliveira et al., 2021	42	Incidental	Words on the screen	Neutral	All	Tobii T60	50 HZ	Blinks and drifts removed, linear interpolation of	0–250 ms during the trial	250–3750 ms	Baseline corrected mean

									missing data, smoothing of data			dilation (variation)
12.	Otero et al., 2011 Experiment 1	45	Not stated	Words on the screen	Neutral	All	EyeLink II	250 HZ	Not stated	0–250 ms during the trial	250–2000 ms	PDR
13.	Otero et al., 2011 Experiment 2 – shallow encoding	31	Intentional	Words by audio	Neutral	All	EyeLink II	250 HZ	Not stated	0–250 ms during the trial	250–3500 ms	PDR
14	Otero et al., 2011 Experiment 2 – deep encoding	31	Intentional	Words by audio	Neutral	All	EyeLink II	250 HZ	Not stated	0–250 ms during the trial	250–3500 ms	PDR
15.	Otero et al., 2011 Experiment 3	33	Not stated	Words on the screen	Neutral	All	EyeLink II	250 HZ	Not stated	0–250 ms during the trial	250–2000 ms	PDR
16.	Taikh & Bodner, 2022	33	Intentional	Words on the screen	Neutral	Correct	EyeLink II	1000 HZ	Blinks and saccades removed	0–250 ms during the trial	250–4000 ms	Baseline corrected mean dilation
17.	Võ et al., 2008	19	Intentional	Words on the screen	Negative, neutral, positive	Correct	iView X Hi-Speed eye tracker	200 HZ	Blinks and artifacts removed	0–200 ms before trial	0–1800 ms	Baseline corrected mean dilation

Note: The table depicts each article in alphabetical order, their respective number of participants, memory type, stimuli used, trials, eye-tracker type with sampling rates, baseline pupil sizes, the window of the pupil response, and the measure of the old-new difference. Descriptions of each moderator are included in the main text.

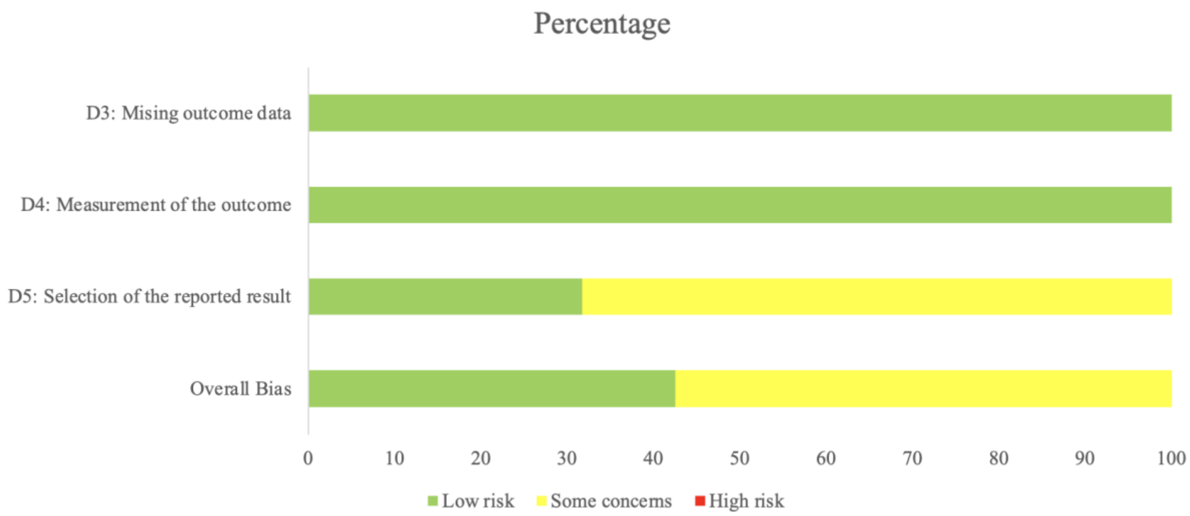
All experiments included in the meta-analysis featured enough statistical information to calculate Cohen's *d*. Out of eight additional authors contacted, one author provided the necessary raw data for the calculation of means and standard deviations, and statistical information for one experiment was found on the OSF. We were able to calculate Cohen's *d* from *F* statistics in four experiments (23.5%), *T* statistics in nine experiments (53%), and means and standard deviations in four experiments (23.5%). Across all experiments, we observed a multitude of approaches to preparing the calculation of the pupil old/new effect. One parameter in particular that might have affected the calculation was the baseline pupil size, which was needed to establish the pupil size change related to old/new stimuli presentation. Some researchers take the baseline pupil size during the presentation of a fixation cross before the stimulus presentation, while others opt to measure the baseline pupil size during the stimulus presentation. Eight out of 17 experiments (47%) used the baseline pupil size from the time window of 1 s *before* the trial, and the remaining 53% of experiments used an interval of 1 s *during* the trial. The time window of the pupil response indicates the period during which the researchers tracked the pupil change from the baseline as a response to a presented stimulus. Seven experiments tracked the pupil change during the full time of the stimulus presentation (stimulus-locked pupil change tracking used in 41% of studies), while the rest chose different approaches. After analyzing the selected literature for the pupil old/new effect measurements, we detected three general categories for determining how old and new pupil size values were compared. The first was the comparison of baseline corrected maximal/peak pupil dilations (used in four experiments, or 23.6%), which was calculated as the subtraction of the baseline pupil size from the maximal or peak pupil size during the recognition period. The second one included comparing baseline corrected mean dilation (implemented in eight experiments, or 47%), which was calculated as the subtraction of the baseline pupil size from the average pupil size during the recollection period. The third category consisted of a comparison of the pupil dilation ratio (PDR; implemented in five experiments, or 29.4%) and was calculated as a proportion of participants' maximum pupil size during the recognition period to participants' maximum pupil size during the baseline period.

Researchers also exhibited various ways of preparing and preprocessing data for statistical analyses. These involved the removal of blinks, removal of signal losses, noise and saccades (11 experiments, or 64.7%), filling in the missing data with linear or spline interpolation (nine experiments, or 52.9%), and, finally, down-sampling the data (six experiments, 35.3%), while two articles with five experiments did not state any information about the method of preprocessing the pupil data.

Risk of Bias Assessment

Figure 2

Risk of Bias Assessment in the Chosen Articles



Note: Domain 3 (D3): missing outcome data. Domain 4 (D4): measurement of the outcome. Domain 5 (D5): selection of the reported result. Overall bias displays the worst outcome in any evaluated domain.

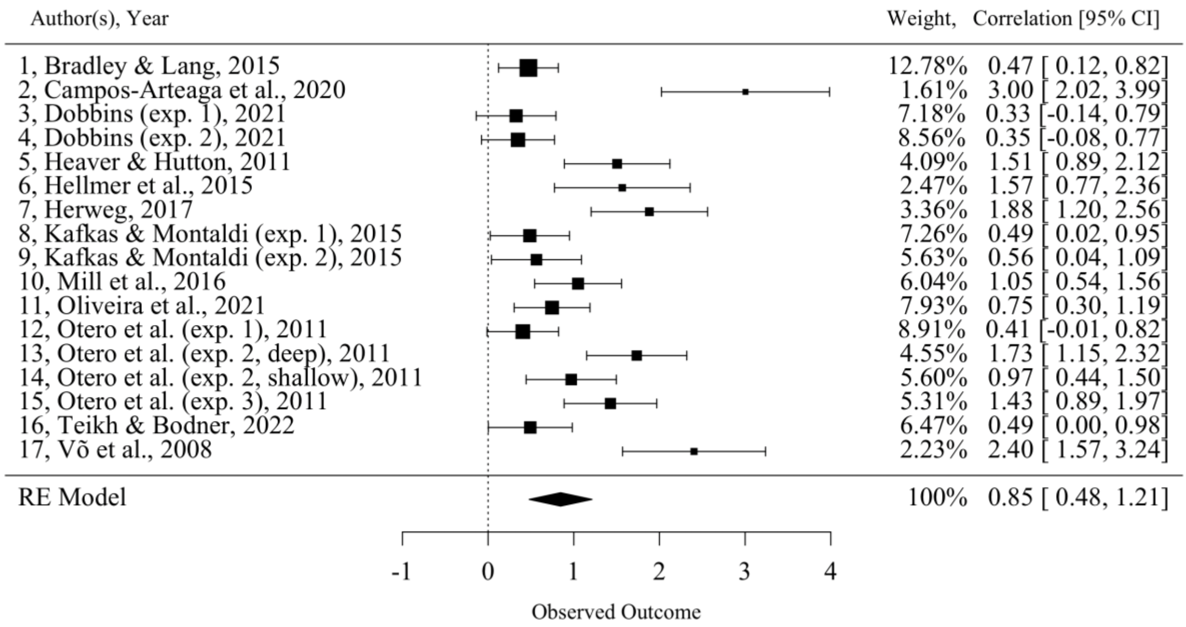
The risk of bias across the chosen articles was evaluated using the Cochrane Risk of Bias 2 tool (Sterne et al., 2019). The assessment was conducted by two independent reviewers; the reported result was the average of their individual assessments. The results are presented in Figure 2, where the evaluation of the domains D3, D4, and D5 is shown. For D3 (missing outcome data), the risk was graded as “low.” Indeed, no measured outcome values were lost during the experiments, and participants who failed or withdrew from the task were excluded from the analysis. Additionally, any missing pupil data were appropriately extrapolated. Similarly, for D4 (measurement of the outcome), the risk was graded as “low.” All articles reported or featured appropriate analyses of the outcomes. However, for D5 (selection of the reported result), articles raised “some concerns,” as no article explicitly reported the intended prespecified plan for the statistical analysis of the pupil old/new effect. Despite the fact that the collective assessment across chosen articles yielded overall “some concerns” in the risk of bias assessment, it is largely due to the absence of standardized approach of reporting the predefined analyses when it comes to investigation of the pupil old/new effect. To our advantage, no article showed “high” risk of bias, which further confirms the validity of selected articles.

Meta-Analytical Results

Altogether, we were able to compute 17 effect sizes. The forest plot (Figure 3) shows the overall pooled effect size across all the chosen articles to be d = 0.85, with a 95% CI [0.48, 1.21], indicating a strong difference between pupil size for old and new stimuli. Thus, the pupil size for old stimuli was visibly larger than that for new stimuli across all 17 experiments.

Figure 3

Forest Plot of Experiments Included in the Meta-Analysis of the Pupil Old/New Effect



Note: Each experiment is represented in alphabetical order on the left side of the plot. Each square represents the individual effect size of the experiment, with its respective confidence interval [95% CI]. Percentages on the right show the weight of each experiment, which consisted of the individual contribution to the overall effect size based on the sample size and standard error. Larger squares indicate larger effect sizes. The overall effect is represented by the rhombus at the bottom of the plot.

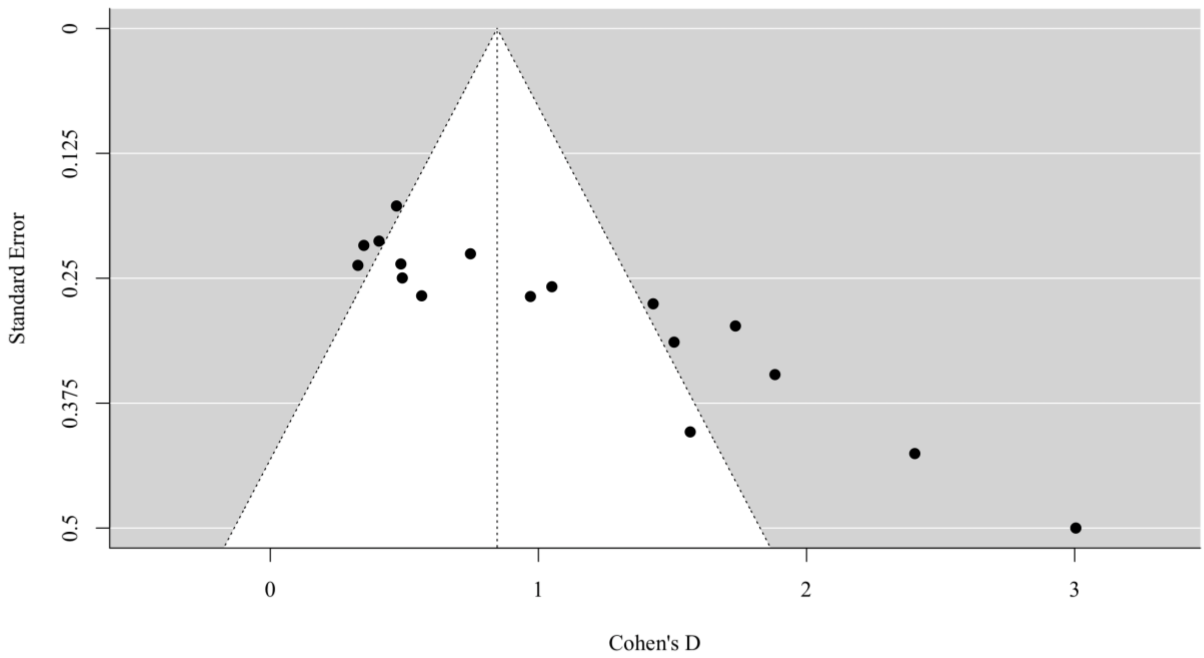
To further investigate the heterogeneity of the articles included in the meta-analysis, we generated a funnel plot (see Figure 4), which visually represents the distribution of effect sizes in relation to their respective standard errors. The funnel plot showed significant asymmetry in effect sizes, which can be explained by two articles with small sample sizes that drove the effects higher.

Egger's Test of the Intercept

We ran Egger's test of the intercept (Egger et al., 1997) to control for funnel plot asymmetry and to check for small-study effects. The results of the Egger's test indicated that the data in the funnel plot were indeed asymmetrical (t (15) = 7.35, p < .001) due to small-study effects.

Figure 4

Funnel Plot of Cohen's d Effect Size Versus Experimental Standard Errors



Selection Model Test

We ran the three-parameter selection model test to determine the extent of the publication bias's influence on the meta-analytical results. The estimated amount of total heterogeneity (*tau^2* = 0.49, *SE* = 0.26) indicated the presence of variability among the effect sizes of the experiments included in the analysis. The likelihood ratio test (*LRT* (1) = 38.25, *p* < .001) showed highly significant results, which indicates a notable amount of heterogeneity among the effect sizes. The results of the three-parameter selection model's results revealed the estimate of the true average effect size to be *d* = 0.87 (*SE* = 0.31, *p* = .005). The plausible range of the effect sizes varied from small to large, with a 95% CI [0.26–1.48]. Interestingly, this estimate was consistent with the pooled effect size that we obtained previously (standard calculation *d* = 0.85, 95% CI [0.48, 1.21]); however, it is important to note that the range was wider, and the three-parameter selection model potentially offers a broader perspective on where the true effect size lay. Overall, this result does not indicate that our meta-analysis was substantially biased by a lower selection probability of non-significant results. This finding is corroborated by the insignificant test of selection model parameters (*LRT* (1) = 0.93, *p* = .340), and there is no strong evidence to suggest significant publication bias.

Given that there was no strong evidence for significant publication bias, but overall heterogeneity was high, we conducted sensitivity analyses of the moderators that might have driven the asymmetry in the effect size distribution. We included the following moderators in the analysis: number of participants, memory measure used in the study design (incidental/intentional), stimuli type (pictures/words), trials chosen for study results (all/correct), eye-tracking sampling rate, baseline measure of the pupil size, time window of the pupil response (as exact time the pupil was tracked for diameter changes), and pupil measure used to identify the pupil old/new effect (baseline corrected peak dilation/baseline corrected mean dilation/PDR; see Table 1).

The combined effect of all moderators included in the analysis was significant (*QM* (8) = 45.9, $p < .001$), indicating that some of the moderators had an impact on the overall effect size. The test for residual heterogeneity showed insignificant results (*QE* (4) = 3.74, $p = 0.4$). A significant effect was associated with the moderator *memory type* (the coefficient was positive and significant [*estimate* = 0.72, $p = 0.028$]), which suggested that intentional memory had a higher effect size compared to incidental memory. Analysis of the *time window of the pupil response* (*estimate* = –0.0003, $p = 0.032$) suggested that an increase in the time window was associated with a decrease in the effect size. In looking at the *sampling rate of the eye tracker* (*estimate* = –0.0015, $p = 0.01$), it was revealed that as the sampling rate increased, the effect size decreased. The results of the *pupil measure* (*estimate* = –1.983, $p = 0.045$) indicated that the baseline corrected peak dilation had a higher effect size compared to the baseline corrected mean dilation and PDR. Finally, a positive trend with a *p*-value close to significance was associated with the *stimulus type* moderator (see Table 2 below).

Table 2.

Statistical Model Results

	Estimate	se	zval	pval	ci.lb	ci.ub	
intercept	3.3739	0.7165	4.7091	<.001	1.9697	4.7781	***
n	-0.0092	0.007	-1.3169	0.1879	-0.0229	0.0045	
trials	0.4722	0.3831	1.2325	0.2177	-0.2787	1.223	
memory type	0.7208	0.3272	2.203	0.0276	0.0276	1.3621	*
time window	-0.0003	0.0001	-2.1468	0.0318	-0.0005	-0.0000	*
sampling rate	-0.0015	0.0006	-2.561	0.0104	-0.0027	-0.0004	*
pupil measure	-1.9832	0.9887	-2.0059	0.0449	-3.9209	-0.0454	*
baseline measure	-0.0001	0.0003	0.2241	0.8227	-0.0007	0.0006	
stimuli type	-1.2544	0.723	-1.735	-0.0827	-2.6713	0.1626	.

Note: N = 14 experiments. Out of the initial 17 experiments, we excluded two experiments due to missing information with regards to the memory type question, as well as one experiment that involved both pictures and words as stimuli type for precise calculation. The table depicts the number of participants, trials, memory type, time window of the pupil response, sampling rate of the eye tracker, pupil measure, baseline measure, and stimuli-type moderators of the meta-analytical results. The "Estimate" column shows the estimated effect size of each moderator; "se" represents the standard error of the estimate; "zval" represents the z-value, indicating how many standard errors the estimate is from zero; "pval" represents the associated p-value; and "ci.lb" and "ci.ub" represent the lower and upper bounds of the confidence interval for the estimate. The variables "n," "time window," "sampling rate," and "baseline measure" are coded as continuous, whereas the variables "trials," "memory type," and "stimuli type" are coded as factors with two levels (all/correct, incidental/intentional, and pictures/words, respectively). Finally, "pupil measure" is coded as a factor variable with three levels (baseline corrected peak dilation, baseline corrected mean dilation, PDR). · $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

To make sure we did not miss any potential moderation, we additionally ran all analyses with only one moderator at the time included in the model. The results were consistent with regard to the *time window of the pupil response*, which remained significant (*QM* (1) = 15.23, $p < .001$, *tau^2* = 0.07 (*SE* = 0.06)), accounting for 72.82% of the heterogeneity. The results suggest that an increase in the time used to track the pupil response was associated with a decrease in the effect size. However, regarding the impact of the *number of participants*, the analysis showed a significant effect on the overall effect size (*QM* (1) = 7.1, $p = 0.008$), which explained 40.89% of the observed heterogeneity. The results suggest that as the number of participants in the experiment increased, the size of the effect tended to decrease. The analysis of *trials chosen for the study results (correct/all)* also showed a significant result (*QM* (1) = 5.04, *p* = .025, *tau^2* = 0.15 (*SE* = 0.1)), with 42.97% of the heterogeneity accounted for. This result suggests that the effect was stronger when the analyses were carried out with all trials (including wrong answers) compared to correct only trials. When testing the *chosen memory type (incidental/intentional)* moderator, our analysis did not reveal significant results (*QM* (1) = 0.62, *p* = 0.43, *tau^2* = 0.29 (*SE* = 0.15). Similarly, the *sampling rate* moderator analysis did not reveal any significant results (*QM* (1) = 1.04, *p* = 0.31, *tau^2* = 0.29 (*SE* = 0.15)), nor did the *stimuli type* moderator analysis (*QM* (1) = 1.44, *p* = 0.23, *tau^2* = 0.265 (*SE* = 0.145).

Outlier Analysis

To control for the quality of the articles included in the meta-analysis, we performed an outlier analysis. Since a study's sample size and standard error are closely related (Harrer, 2022), larger standard errors of effect sizes result in wider CI and increase the chance that the effect is not statistically significant. Experiments were defined as outliers when their 95% CI was outside the 95% CI of the overall pooled effect. From this analysis, two outliers emerged, which included the experiments by Campos-Arteaga et al. (2020) and Võ et al. (2008; see Table 1). Both experiments showed large effect sizes (*d* = 3 and d = 2.4, respectively) with relatively small sample sizes (*n* = 17 and *n* = 19, respectively) and large standard errors (*SE* = 0.50 and *SE* = 0.43, respectively). Upon removing these outliers, the recalculated meta-analytical results (*n* = 15) showed a slightly increased overall effect size and a narrower range (*d* = 0.89, CI [0.62, 1.15]) than the initial findings (*d* = 0.85, CI [0.48, 1.21]), which potentially indicates a more accurate and precise estimation of the effect size.

Discussion

The present meta-analysis has focused on systematizing and analyzing the literature that investigates the pupil old/new effect, which manifests in a larger pupil dilation in response to old stimuli versus new stimuli as a measure of memory. Despite the various mentions of larger pupil dilation in response to old versus new stimuli, we found only a small number of articles that investigated the pupil old/new effect and contained enough information about the calculation of the effect. Across the 293 articles identified in the initial search, we found 12 articles with 17 experiments that fit our preidentified inclusion criteria. The analysis of the experiments was performed according to the PRISMA guidelines (Page et al., 2021).

The risk of bias analysis, carried out by two independent researchers, revealed a "low" to "some concerns" risk of bias across all the chosen articles, which is linked to the absence of standardized practices of prespecifying the intended analyses regarding the pupil old/new effect, however, no article reported "high" risk of bias, which justifies the validity of the statistical analyses performed in this meta-analysis. Notably, our findings revealed a large positive effect (Cohen's $d = 0.85$) for the pupil old/new effect. This effect was consistently positive across all articles included in the analysis and ranged from 0.48 (indicating a medium effect) to 1.21 (indicating a very large effect). During our analysis, we detected two outliers in the pool of chosen articles. Upon removal, the effect size slightly increased and exhibited a narrower range ($d = 0.89$, CI [0.62, 1.15]). This adjustment refined the overall estimation of the effect.

Furthermore, after fitting the funnel plot and subsequently applying Egger's test of the intercept, we identified asymmetry resulting from small-study effects. To assess the variability in the effect sizes, we performed a three-parameter selection model test, which indicated existent heterogeneity in the effect sizes; however, it did not provide strong evidence for significant publication bias.

To further examine and identify the source of variability, we investigated potential moderators that could have enhanced or diminished the pupil old/new effect. First, we integrated all moderators into one model, which is crucial for studying their collective impact on the effect size. Among the characteristics tested (see Table 1), four moderators (*memory type*, *time window of the pupil response*, *sampling rate of the eye tracker*, and the *pupil measure*) showed significant results. Further individual analyses of each moderator in separate models confirmed the four moderators, and showed that two additional moderators (the *number of participants and trials chosen for study results* and the *time window of the pupil response*) also emerged as significant. The significant moderators indicated that the pupil old/new effect appeared more pronounced with intentional (vs. incidental) memory, with shorter (vs. longer) time windows of the pupil response, with a higher sampling rate of the eye tracker, with PDR (vs. baseline corrected mean dilation or baseline corrected peak dilation) as the pupil measure, with smaller sample size, and with all trials (including incorrect answers). This last point aligns with previous findings suggesting that the pupil old/new effect presents itself regardless of correct or incorrect answers on the memory recognition test (e.g., Montefinese et al., 2013; Otero et al., 2011).

Limitations and Future Directions

The findings of this meta-analysis may reflect some limitations. First, a more extensive exploration of moderators is needed. While we have examined several moderators, such as the memory/stimuli/trial type, number of participants, baseline measure of the pupil response, time window of the pupil response, and the pupil old/new effect measure, some moderators were represented by only one or two experiments, making it challenging to draw any conclusions about their impact. For example, research on the pupil old/new effect could largely benefit from an investigation into the nature of stimuli (e.g., complexity of the stimuli and emotional impact) presented to participants during the memory task, with the purpose of exploring the consistency of the pupil old/new effect across different stimuli types. Investigating temporal characteristics, like the difference between initial (500–1000 ms after the onset of stimuli) and late pupil dilation windows (1000–3000 ms after the onset of stimuli; Bradley & Lang, 2015) could improve our understanding of concurrent cognitive processes and perhaps help us to distinguish between them.

In future research, employing larger sample sizes (*n* > 25), based on the availability of laboratory resources, seems crucial for the investigation of the pupil old/new effect. Our outlier analysis suggests a need for larger sample sizes to prevent potential inflation of the effect size due to chance. Furthermore, researchers should justify whether they include or exclude accuracy in the statistical models (taking only correct trials or both correct and incorrect trials for the analysis), depending on the research question.

Moreover, research on the pupil old/new effect can greatly benefit from exploring different populations. Some populations have already been tested; for example, the effect was found to be present in amnesic patients by Laeng et al. (2007) and absent in the autism spectrum disorder (ASD) group in the work of Ring et al. (2020). The effect has also been tested in seven-month-old children (Hellmer et al., 2015), although the temporal dynamics here differed from those associated with studies on adults, as children required two sessions of exposure to the old items to show the pupil old/new effect. Nevertheless, a clear difference between pupil size for old and new stimuli was present in children of such a young age. An in-depth investigation of different populations is needed to generate a more robust database. Further investigation into the effect in clinical populations, older adults, or individuals in whom one sensory modality is not available (e.g., in people who cannot speak, hear, or see) can also reveal interesting insights into memory-related deficits and limitations associated with the pupil response to old and new information.

Conclusion

This meta-analysis was performed to provide a comprehensive analysis of existing research on the pupil old/new effect as a measure of memory. A consistently positive and large effect size (Cohen's $d = 0.85$) was evident across all the included articles, despite high heterogeneity and significant moderators. The consistency of this effect across various experimental designs and analytical approaches highlights its robustness. Taken together, these findings suggest that the pupil old/new effect can be further used in various domains of research as an implicit measure of memory. Nevertheless, more research is needed on potential moderators and different populations; while the relationship between pupil dilation and the recognition associated with the old/new effect is an intriguing phenomenon, the exact mechanisms underlying it remain undetermined.

Acknowledgments

We would like to thank the authors who provided additional information about their articles. We would also like to thank Sarah Schnyder and Marine Beney for their help with the literature search and the risk of bias analysis.

Disclosure Statement

The authors report no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in OSF at https://osf.io/5pgcj/?view_only=51e59437d86e4cc8b1bff4e582f25b8c.

ORCID

Corinna Martarelli: https://orcid.org/0000-0001-9160-793X

Alexandra Lapteva https://orcid.org/0000-0002-8502-1294

References

Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., Van De Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14*(4), 1–1. https://doi.org/10.1167/14.4.1

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Barendregt, J. J., Doi, S. A., Lee, Y. Y., Norman, R. E., & Vos, T. (2013). Meta-analysis of prevalence. *Journal of Epidemiology and Community Health*, *67*(11), 974–978. https://doi.org/10.1136/jech-2013-203104

Beatty, J. (1982). *Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources*. *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Binda, P., Pereverzeva, M., & Murray, S. O. (2014). Pupil size reflects the focus of feature-based attention. *Journal of Neurophysiology*, *112*(12), 3046–3052. https://doi.org/10.1152/jn.00502.2014

Bradley, M. M., & Lang, P. J. (2015). Memory, emotion, and pupil diameter: Repetition of natural scenes: Memory, emotion, and pupil diameter. *Psychophysiology*, *52*(9), 1186–1193. https://doi.org/10.1111/psyp.12442

Brocher, A., & Graf, T. (2016). Pupil old/new effects reflect stimulus encoding and decoding in short-term memory. *Psychophysiology*, *53*(12), 1823–1835. https://doi.org/10.1111/psyp.12770

Brown, M. W., & Xiang, J.-Z. (1998). Recognition memory: Neuronal substrates of the judgement of prior occurrence. *Progress in Neurobiology*, *55*(2), 149–189. https://doi.org/10.1016/S0301-0082(98)00002-1

Campos-Arteaga, G., Forcato, C., Wainstein, G., Lagos, R., Palacios-García, I., Artigas, C., Morales, R., Pedreira, M. E., & Rodríguez, E. (2020). Differential neurophysiological correlates of retrieval of consolidated and reconsolidated memories in humans: An ERP and pupillometry study. *Neurobiology of Learning and Memory*, *174*, 107279. https://doi.org/10.1016/j.nlm.2020.107279

Coburn, K. M., & Vevea, J. L. (2015). Publication bias as a function of study characteristics. *Psychological Methods*, *20*(3), 310–330. https://doi.org/10.1037/met0000046

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research*, *15*(2), 191–205. https://doi.org/10.1016/S0926-6410(02)00192-1

Dobbins, I. G. (2021). Pupil dilation signals recognition salience. *Psychonomic Bulletin & Review*, *28*(2), 565–573. https://doi.org/10.3758/s13423-020-01866-w

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*(2), 170–177. https://doi.org/10.1037/1082-989X.1.2.170

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

Fanelli, D. (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLoS ONE*, *5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Harrer, M. (2022). *Doing meta-analysis with R: A hands-on guide* (First edition). CRC Press.

Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory*, *19*(4), 398–405. https://doi.org/10.1080/09658211.2011.575788

Hedges, L. V. (1984). Estimation of Effect Size under Nonrandom Sampling: The Effects of Censoring Studies Yielding Statistically Insignificant Mean Differences. *Journal of Educational Statistics*, *9*(1), 61–85. https://doi.org/10.2307/1164832

Hellmer, K., Söderlund, H., & Gredebäck, G. (2018). The eye of the retriever: Developing episodic memory mechanisms in preverbal infants assessed through pupil dilation. *Developmental Science*, *21*(2), e12520. https://doi.org/10.1111/desc.12520

Herweg, N. A., Sommer, T., & Bunzeck, N. (2018). Retrieval Demands Adaptively Change Striatal Old/New Signals and Boost Subsequent Long-Term Memory. *The Journal of Neuroscience*, *38*(3), 745–754. https://doi.org/10.1523/JNEUROSCI.1315-17.2017

Kafkas, A., & Montaldi, D. (2011). Recognition Memory Strength is Predicted by Pupillary Responses at Encoding While Fixation Patterns Distinguish Recollection from Familiarity. *Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989. https://doi.org/10.1080/17470218.2011.588335

Kafkas, A., & Montaldi, D. (2015). The pupillary response discriminates between subjective and objective familiarity and novelty. *Psychophysiology*, *52*(10), 1305–1316. https://doi.org/10.1111/psyp.12471

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., Fabian, V., & Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, *8*(1), 4949. https://doi.org/10.1038/s41598-018-23197-6

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. https://doi.org/10.3758/BF03210951

Lüdecke, D. (2018). *esc: Effect Size Computation for Meta Analysis* (0.4.1) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.1249218

Maillet, D., & Schacter, D. L. (2016). When the mind wanders: Distinguishing stimulus-dependent from stimulus-independent thoughts during incidental encoding in young and older adults. *Psychology and Aging*, *31*(4), 370–379. https://doi.org/10.1037/pag0000099

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Mill, R. D., O'Connor, A. R., & Dobbins, I. G. (2016). Pupil dilation during recognition memory: Isolating unexpected recognition from judgment uncertainty. *Cognition*, *154*, 81–94. https://doi.org/10.1016/j.cognition.2016.05.018

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The "subjective" pupil old/new effect: Is the truth plain to see? *International Journal of Psychophysiology*, *89*(1), 48–56. https://doi.org/10.1016/j.ijpsycho.2013.05.001

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), 11. https://doi.org/10.1167/13.2.11

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, *8*(4), 434–447. https://doi.org/10.1037/1082-989X.8.4.434

Oliveira, J., Fernandes, M., Rosa, P. J., & Gamito, P. (2021). Is Pupil Activity Associated With the Strength of Memory Signal for Words in a Continuous Recognition Memory Paradigm? *Frontiers in Psychology, 12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.686183

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *48*(10), 1346–1353. https://doi.org/10.1111/j.1469-8986.2011.01217.x

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. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, *83*(1), 56–64. https://doi.org/10.1016/j.ijpsycho.2011.10.002

R Core Team. (n.d.). [Computer software].

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis. *Journal of Clinical Epidemiology*, *53*(11), 1119–1129. https://doi.org/10.1016/S0895-4356(00)00242-0

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, l4898. https://doi.org/10.1136/bmj.l4898

Taikh, A., & Bodner, G. E. (2022). Pupil dilation during recognition reflects the subjective recollection/familiarity experience at test rather than the level of processing at encoding. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *76*(3), 186–192. https://doi.org/10.1037/cep0000283

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(6), 607–617. https://doi.org/10.1037/0278-7393.5.6.607

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

Võ, M. L.-H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, *45*(1), 130–140. https://doi.org/10.1111/j.1469-8986.2007.00606.x

Wang, C.-A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, *33*, 134–140. https://doi.org/10.1016/j.conb.2015.03.018