**Acute slowing of reaction times already at low doses**

**of psilocybin on attention and executive functioning: a systematic review and meta analysis**

Yousefi, P.a, , O”Higgins, F. J.b,d/ Lietz, M.P. b,c \* / O”Higgins, F. J.b,d, Rippe, R.C.A.e, van Elk Ma,f,, Enriquez-Geppert, S.a,g \*

aDepartment of Psychology, Leiden University, the Netherlands

bDepartment of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, the Netherlands

a XXX, Ireland

e Institute of Education and Child Studies, Leiden University, the Netherlands

f Amsterdam, the Netherlands

g Department of Biomedical Sciences of Cells & Systems, University Medical Center Groningen, Groningen, the Netherlands

**Running head:** CCCC

**Keywords:** psilocybin, acute effects, cognition, attention, executive function

**Manuscript type**: systematic review & Meta-analysis

**Word count:** CCCCC

**Potential journals:** Psychopharmacology IIII (3.424), J of Psychoactive Drugs, Neuropsychopharmacology (2.536)

**Correspondence:**  
Stefanie Enriquez-Geppert  
Department of Clinical and Developmental Neuropsychology  
Faculty of Behavioural and Social Sciences  
University of Groningen  
Grote Kruisstraat 2/1  
9712 TS Groningen  
The Netherlands  
Email: [s.enriquez.geppert@rug.nl](mailto:s.enriquez.geppert@rug.nl)

**ABSTRACT**

**Introduction:** xxx

**Method:** xxx

**Results:** Txxx

**Discussion and conclusions:** xxx

**Applications:** xxx

1. **INTRODUCTION**

* the most commonly used psychedelic in clinical studies
* why cognition…. thalamic gating theory, deficits are reported but is it more transient attentive problems? lack of attentional focus?
* Cognitive recovery or enhancement after psychedelic? Are there long-term measures? or only acute? unlocking of full potential in der akuten Phase möglich? Lack of depression and improvement of cognition?
* Psychedelic und cogntive enhancement?

Psilocybin (and its metabolite),, has gained increasing interest in recent years due to its potential therapeutic effects on various neuropsychiatric disorders, including depression (Carhart-Harris et al., 2021; Li et al., 2022; McCartney et al., 2022; Więckiewicz et al., 2021), anxiety (Griffiths et al., 2016; Ross et al., 2016) and substance use disorder (Bogenschutz et al., 2022, Johnson 2016…). While the effects of psilocybin on emotions and psychological functioning have been more extensively studied (Barrett et al., 2020; Basedow et al., 2021; Irizarry et al., 2022; Nutt & Carhart-Harris, 2021), its impact on cognition remains a topic of ongoing research (Bonnieux et al., 2023; Sayalı & Barrett, 2023). Addressing this gap, this meta-analysis aims to examine the effects of psilocybin on cognition, specifically focusing on its influence on various cognitive domains such as memory, attention, and executive function.

**Effects of psilocybin on cognition**

Studies have explored the impact of psilocybin on cognitive processes such as attention (Cavanna et al., 2022), working memory (Barrett et al., 2018), inhibition (Doss et al., 2021; Kometer et al., 2012; Marschall et al., 2021), and creativity (Bonnieux et al., 2023; Sayalı & Barrett, 2023; Mason et al., 2019).

**Mechanisms of action**

Understanding the cognitive effects of psilocybin requires insight into its pharmacological mechanisms. Psilocybin metabolises into the psychoactive psilocin, (Geiger et al., 2018), both of which are found in a range of hallucinogenic mushrooms. This metabolite primarily targets serotonin 5-HT2A receptors, and secondarily to other serotonin, histamine, and adrenergic receptors (Halberstadt & Geyer, 2011; Lewis et al., 2020; Waters, 2021). These biochemical interactions lead to a range of well-established cognitive, emotional, and perceptual changes in human behavior.

* are there differences reagding actue and long term effects\_
* What do we know about microdosing and a dose-response effects?

Prefrontal cortex has a lot of 5ht2a receptors.

5ht2a activity results in excitation ()

mention role of microglia (VanderZwaag et al., 2023).

**continue here**

On a larger scale, the effect of psilocybin has been argued to be related to the accurary vs speed trade off, and to cognitive flexibility vs stability( SOURCEs from neurdomodulation coutrse)

* Write about: the accuracy speed trade off + cognitive flexibility vs stability (Sayalı & Barrett, 2023)
  + Write more extensively about Bonnieux et al., 2023 and Sayalı & Barrett, 2023 and what the newest ideas are about the effects of psilo on cognition.

Despite the growing body of research on the effects of psilocybin on cognition, there are still unclarity about the overall effects and - also because of- methodological challenges in this field (Hendy, 2018; Van Elk & Fried, 2023). Besides the problems with placebo groups and expectancy effects, the variability in dosages and administration protocols across studies makes it difficult to compare and generalize the findings. Additionally, the use of different cognitive assessment measures for different domains of cognition and the lack of standardized protocols further complicate the interpretation of results. Furthermore, the effects of psilocybin on cognitive functions may be influenced by individual factors such as, set and setting (Studerus et al., 2012; Viktorin et al., 2022). Given the mixed findings and methodological challenges in the existing literature, a comprehensive and critical appraisal of the effects of psilocybin on cognition is needed.

The present (systematic review and) meta-analysis is designed to fill this research void by evaluating the effects of psilocybin on cognition, with a specific focus on executive functions as delineated by Miyake and Friedman's model (Miyake et al., 2000). According to this model, executive functions comprise updating (working memory), inhibition, and shifting (cognitive flexibility). Notably, we also incorporate conflict monitoring as a distinct and vital aspect of executive function, as suggested by Enriquez-Geppert et al. (2010). Our study aims to determine the pooled effects of psilocybin on these subdomains of cognition and to examine, cognitive category, dose and measurement time point as potential moderators. To our knowledge, this is the first meta-analysis to undertake such a comprehensive evaluation.

**2. METHODS**

2.1 Literature Search and Inclusion Criteria

A comprehensive systematic review was conducted by searching multiple electronic databases, including PubMed,PsychInfo, Web of Science, and Cochrane) to identify empirical articles on psilocybin and executive functions using the key search terms *cognition* or *cognitive function\** or *executive function\** or *cognitive control* or *inhibition* or *memory* *updating* or *conflict monitoring* or *task switching* or *set-shifting,* combined with one of the following terms: *psychedelic\** or *hallucinogen\** or *psiloc\** or *psychotomimetic* or *entheog\** or *\*shrooms\**. We searched for articles published until January 2023. In July 2023, the same search strategy was applied one more time to update the database, resulting in the inclusion of one more study.

2.1.1 Inclusion and exclusion criteria

To meet the inclusion criteria, papers had to satisfy the following requirements: Included studies had to report the measurement of at least one of the following cognitive domains under the influence of psilocybin: (a) working memory (updating) (b) conflict monitoring c) response inhibition (d) cognitive flexibility (e) attention. We did not differentiate between acute or long-term effects, nor in which form or dosage psilocybin was administered. Studies that reported the measure of more than one cognitive subdomain separately were treated as independent studies.

Exclusion criteria encompassed studies that (1) were not written in the English language, (2) did not involve the administration of any form of psilocybin, (3) utilised an incorrect study design (animal models, or lack of relevant executive functions), (4) were of an incorrect publication type (background article, reviews, meta-analysis, dissertation), (5) were inaccessible.

The software Rayyan (Ouzzani et al., 2016) was used for screening abstracts. The automated tool of the software was used for the detection of duplication. Four independent reviewers were responsible for screening abstracts of each report, working independently to ensure accuracy and consistency. For the exclusion of a study, the assessment of only one reviewer was sufficient. However, for the inclusion of the study, at least two reviewers had to include the study. The studies were excluded in a hierarchical manner, meaning that if a study was excluded because of one reason, it was not assessed for other reasons and was excluded immediately.

2.2 Data extraction

To systematically collect data, five authors investigated the full-text articles of the selected studies. In instances of missing data for calculating effect sizes, the corresponding author of the respective study was contacted for clarification. Multiple outcome domains were targeted: cognitive function, specific cognitive measures, sample size for both experimental and control groups (N), dependent variables for each cognitive measure, dosage, measurement time points, and either pre-calculated effect sizes or the raw data required to calculate them. In terms of other variables, data were also collected on participant and intervention characteristics (see supplementary material). The individual results were visualised in an excel sheet for comparison and subsequently included in the manuscript (Table 3).

A comprehensive strategy was employed, seeking all results compatible with each outcome domain, across all available measures, time points, and analyses within each study. We categorised the dosages of the included studies according to the following categories: 1-5 mg micro, 6-19 mg: low, 20-30 mg: medium, >30 mg: high (all per 70kg).

For the meta analytical procedures, a decision was made to concentrate on specific cognitive outcome measures of reaction time (RTs) and accuracy (ACC), given their prominence and relevance in the included studies. Effect measures for these outcome measures were reported in mean and median differences, which were calculated into standardised values and used as the basis for the calculation of effect sizes for each study. The scripts for the calculation of the effect sizes for each study and the extracted data used for these calculations can be found in the supplementary material. For the extraction of data from graphical representations in individual studies, the software WebPlotDigitizer (Rohatgi, 2022) was utilised**.**

2.3 Risk of bias assessment

Two independent reviewers systematically evaluated the potential for bias using the Cochrane Collaboration's tool for assessing risk of bias (Higgins et al., 2011). To visualize these assessments, we employed the Risk-of-Bias Visualization software (robvis) developed by McGuinness and Higgins (2020).

In addition to this conventional approach, we also meticulously extracted data related to unique sources of bias that are particularly pertinent to psychedelic research (Van Elk & Fried, 2023). These specialised areas of concern included: the efficacy of blinding procedures, the presence of active control groups, participants' prior experience with psychedelics, the existence of a non-intervention study arm, and both participant and researcher expectations. Furthermore, we assessed whether the studies under review had explicitly stated their inclusion and exclusion criteria (see supplementary material). For a comprehensive discussion of these bias threats in psychedelic research, readers are referred to Van Elk and Fried (2023).

2.4 Meta-analysis

To synthesise the collected data, we employed a multilevel meta-analysis approach. This statistical method was chosen to accommodate the complex structure of our data, specifically the presence of multiple effect sizes extracted from single studies. A random-effects model was utilised across all analyses due to the expected variability in study designs, participant characteristics, and psilocybin dosages. Specifically, we initiated our analysis with a multilevel random-effects model, acknowledging the nesting of multiple effect sizes within individual studies. This strategy allowed for the assessment of within-study and between-study variances, essential for our heterogeneous set of studies. Hedges' g was selected as the effect size measure.

Heterogeneity among study results was quantified using the I² statistic and the Q statistic. Substantial heterogeneity was addressed by excluding certain study conditions (e.g., microdosing studies) to achieve a more homogenous set of effect sizes.To assess the potential risk of publication bias, we employed a funnel plot analysis, supplemented by Kendall’s rank correlation test. Additionally, we calculated Rosenthal’s, Rosenberg’s, and Orwin’s fail-safe numbers to estimate the number of missing studies that would render the observed effect size non-significant. To test the robustness of our findings, we conducted sensitivity analyses by comparing results from the multilevel model with those from a non-nested random-effects model. In subsequent analyses, we examined potential moderators (Dosage levels, cognitive functional categories, and timing relative to psilocybin peak effects). Statistical procedures for the meta analysis were conducted using R (R Core Team, 2020), RStudio (Rstudio Team, 2020), and the metafor package (Viechtbauer, 2010).

| **Table 1** |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Included studies an Study IDs* | |  |  |  |
| Study | Study ID | Nr of tasks | Nr of diff dosages | Nr EFs extracted |
| Barrett et al 2018 | 1 | 3 | 3 | 9 |
| Carter et al 2015 | 2 | 2 | 1 | 2 |
| Cavanna et al 2022 | 3 |  |  |  |
| Doss et al 2021 | 4 |  |  |  |
|  | 5 |  |  |  |
| Hasler et al 2004 | 6 |  |  |  |
| Kometer et al 2012 | 7 |  |  |  |
| Quednow et al 2011 | 8 |  |  |  |
| Vollenweider et al 2007 | 9 |  |  |  |
| Wittmann et al 2006 | 10 |  |  |  |
| Marschall et al 2022 | 11 |  |  |  |
| Mallaroni et al 2023 | 12 |  |  |  |

| **Table 2** |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Overview of cognitive measures used in the included studies* | | | | | |
| Domain | | Task | Study IDs | Description | Variables |
| Specific EFs | Working memory (updating) | N-back | (1) | The N-Back task is a measure of working memory (updating). Participants view a sequence of letters and must indicate whether the current letter matches the one presented "n" steps back. | (1)   * Discriminately rate (hit rate [HR] - false alarm rate [FAR]) * Response bias (FAR/[1-accuracy]) * RT |
|
| Spatial Memory Task (SMT) | (12) | The Spatial Memory Test (SMT) is used mainly in psychopharmacological drug research to assess visuospatial memory and reasoning. It has two phases: immediate and delayed recall.  In the immediate recall, participants view six sets of ten black-and-white images on a computer screen, each for 2 seconds with a 1-second gap in between. Subsequently, they must identify the original location of each image as it reappears on the screen. After a 30-minute interval, the delayed recall phase requires participants to recall the locations of the pictures. | (12)   * Total number correct * Mean Reaction Time |
| Spatial Span Test | (10)  (2) | The spatial range test is a measure of visual-spatial working memory performance. Nine white squares are randomly displayed on a black background on a touch screen. On each trial, some of the squares are highlighted at a specific interstimulus interval. Participants indicate the order of the highlighted squares. The trial starts with two squares and increases up to nine. The test ends after three wrong answers or reaching a maximum of 9 squares (Max. score). The number of squares correctly reproduced is the span-length. | (2)   * Span-length * # of errors   (10)  - Span-length |
| Conflict monitoring | Stroop task | (4) (3) (8) | The Stroop task is a measure of conflict monitoring. Typically a colour word is presented in a coulour which can match or not (interference).  Participants are asked to name the ink colour of colour words, which can either be congruent (e.g., the word "red" in red ink) or incongruent (e.g., the word "red" in blue ink). The incongruent condition usually results in longer reaction times and more errors, illustrating the cognitive conflict between word reading and colour naming ("Stroop effect") | (3)  - RT  - ACC  (4)  - Hit Rate -median RT  - Signal detection (discrimination d')  (8)  - RT  - # errors  - Interference-, facilitation-, Stroop effect |
| Stroop task (emotional) | (1) | The emotional Stroop task is a measure of emotional cognitive control. Participants are presented with words that are emotionally salient and neutral words. Participants indicate the colour in which the words are presented. | (1)  - RT  - ACC |
| Inhibition | Go/NoGo | (3) | The Go/NoGo task is a measure of motor inhibition. Participants are required to respond (Go) or withhold their response (NoGo) to specific stimuli, allowing for the measurement of the ability to inhibit prepotent responses. | (3):  - RT  - ACC |
| Go/NoGo (emotional) | (7) (11) | The emotional Go/NoGo task is a measure of emotional cognitive control. Participants see affective stimuli (e.g., positive, negative, or neutral words) and respond only to a specific emotional category (e.g., positive) while withholding a response to another (e.g., negative or neutral). | (7)  - RT  - ACC  - Sequential effects  (11)  - ACC  - RT |
| Multiple EF | Digit-Symbol Substitution Task (DSST) | (1) (12) | The DSST is a measure of processing speed, working memory, and attention. Participants are presented with a key that pairs specific digits with symbols and are asked to quickly fill in a series of boxes with the corresponding symbols for a sequence of digits. | (1)  - total # of attempted trials  - proportion of correct attempted trials  - Accuracy (% correct)  (12)  - Total #Reponses  -Total #Correct  - Mean RT |
| Penn Conditional Exclusion Test (PCET) | (4) | The PCET is a measure of multiple executive functions. The task resembles the Wisconsin Card Sorting Test.  Participants have to identify the common characteristic among a set of items. | (4)  - Preservation errors  - Median RT  - Correct, incorrect, total responses |
| Tower of London (TOL) | (12) | TOL is a measure of multiple executive functions, specif. planning ability. Participants are presented with two sets of different beads: a set representing the desired goal, and a different set that has to be rearranged by the participant. Participants rearrange with a minimal number of moves. This measure was originally developed to test for severe cognitive dysfunctions. | (12)  - Total correct trials  - RT |
| Attention | | Frankfurt Attention Inventory (FAIR) | (6) (9) | FAIR is a paper-pencil measure of sustained and directed attention. Participants identify target stimuli within distractor items within a limited time by spiking at each target and drawing a continuous line below the distractors. | (6)  - Marker Value  - Performance value  - Quality value  - Continuity value  (9)  -performance score (P)  -amount of made decisions /total decisions (Q)  -continuity score (C) |
| Attentional blink task | (3) | The attentional blink task is a measure for allocating attention over time.  Participants view a rapid stream of letters with two digit targets (T1 and T2) inserted. The aim is to identify both digits, and performance is evaluated based on correct detection rates. When T2 appears with a short stimulus onset interval, T2 is often not detected. The task incorporates varying time lags between the targets to assess the duration of the attentional "blink" (Shapiro et al., 1997). | (3)  - Visibility rate of targets (T1,T2) |
| Trail Making Test | (3) | consists of two parts. Trail A requires connecting numbered circles in sequence, while Trail B involves alternating between numbers and letters, with the time taken to complete each part, and total number of errors being recorded. (Reitan & Wolfson, 1995). | (3)  - Time to finish the task  -Number of errors |
| Covert orienting of attention task (COVAT) | (5) | The COVAT assesses attentional control by requiring participants to quickly respond to visual stimuli appearing in different locations on a screen. Cues indicate where the stimulus may appear, but they may be misleading. | (5)  -RT  -validity effect (the difference between invalid and valid reaction times) |
| Multiple Objects Tracking (MOT) | (2) | MOT is a measure of divided visual attention to moving visual stimuli. During the task, participants keep their gaze on a fixation cross and look at a small set of identical stimuli of targets and distractors that start moving randomly on a display. Participants observe these moving objects among distractors and identifying them post-movement | (2)  - The mean % of correct responses  - Mean # of successfully tracked targets  - Discriminability index |

3. Results

3.1 Study selection

The search resulted in a total of 2543 articles, which were screened by title (Figure 1). Articles that did not meet the inclusion criteria were excluded at this point. The remaining articles were uploaded into Ryaan, duplicates were removed, and the abstracts (and if relevant, the full-text) were screened. The screening process was conducted by four independent reviewers. Eventually, a total of 12 studies were suited for the present systematic review. Study Ids 4 and 5 ( included in the literature review) were excluded from the subsequent meta analysis, as the data for the calculation of effect sizes could not be retrieved.

Diagram

Description automatically generated

Figure 1 The hierarchy of exclusion reasons was in the following order: wrong language, wrong drug, wrong design, wrong publication type, not accessible.

3.2 Risk of bias

3.3 Literature review

This literature review delves into the effects of psilocybin on cognitive processes, specifically focusing on attention, working memory, conflict monitoring, inhibition, and cognitive flexibility. The included studies have utilised a variety of cognitive tests to assess these different aspects of attention and cognition. An overview of the tasks and their outcome measures is presented in Table 2 for reference.

**3.3.1 Effects of psilocybin on working memory (updating) (Study IDs 1,2 10, 12)**

Four key studies—Barrett et al. (2018), Carter et al. (2005), Mallaroni et al. (2023), and Wittmann et al. (2006)—considered working memory. The studies pursued different research questions and utilised various assessment tools, including the Letter N-back, Spatial Span Test, and the Spatial Memory Task (Table 2).

Thet study by Barrett et al. (2018; ID = 1), employed a double-blind, placebo-controlled, complete-crossover design to compare the neurocognitive effects of two different types of psychedelics, psilocybin and DXM (an NMDAr antagonist) as a classic and dissociative hallucinogen on cognitive performance. Various doses of psilocybin and a control dose of DXM were administered in 19 participants, followed by cognitive assessment (psychomotor functioning, working and episodic memory, visual perception and executive functions, such as working memory) including a Letter N-back task for working memory updating, conducted at baseline and 180 minutes after compound intake. There was no global cognitive impairment and delirium, but more specific impairments that were drug and dose dependent. For the 2-back condition, the study found that psilocybin led to both a decrease in discriminability and slowed RTs for correct responses. A dose-dependent reduction in response bias was also observed, leading the authors to conclude that psilocybin selectively impairs working memory (Barrett et al., 2018).

The second study by Carter et al. (2005; ID = 2) examined the effects of psilocybin as an experimental approach to assess the involvement of the serotonin receptor in attention and working memory. The study sample consisted of eight participants, both with and without pre-treatment using the 5-HT2A antagonist Ketanserin. Administered in a double-blind, placebo-controlled study, a Spatial Span Test was conducted 120 minutes post-compound intake. The results indicated no significant effect of psilocybin on any of the measured outcomes for working memory, but reduced attentional tracking ability (Carter et al., 2005).

Wittmann et al. (2006; ID = 10) focused on time perception under the effects of psilocybin. As a cognitive control measure, spatial working memory was assessed using the Spatial Span Test. The study was conducted double-blind, placebo-controlled, within-subject design involving 12 participants. Assessments were made at baseline, 100 minutes, and 360 minutes following drug administration. Timing performance was affected for longer time intervals (from 2/2.5 s on). Regarding working memory, dose and time dependent effects on span length were found at acute effects only and reduced the lengths from about 7.5 to 6.5. A 17.5 mg/70 kg dose of psilocybin significantly impaired span length at 100 minutes post-psilocybin-administration, but not after 360 min, in contrast to the 8.05 mg/70 kg dose and placebo (Wittmann et al., 2006).

Lastly, the recent study by Mallaroni et al. (2023; ID = 12) assessed the immediate subjective, cognitive and cardiovascular effects of the novel serotonergic hallucinogen 2C-B (a synthetic analogue of mescaline) in comparison to psilocybin in a double-blind, placebo-controlled within-subject design study. Cognitive domains comprised psychomotor task and executive functions. Mallaroni et al. (2023) reported similar impairments in working memory, as gauged by the Spatial Memory Task (SMT) as well as psychomotor slowing. The study revealed that a 15 mg/70 kg dose of psilocybin, compared to a placebo, was associated with significantly lower scores in both immediate and delayed SMT recall.

In summary, the existing literature presents a nuanced view of the effects of psilocybin on working memory. While Barrett et al. (2018) and Mallaroni et al. (2023) found definitive impairments in working memory following psilocybin administration, the results from Carter et al. (2005) and Wittmann et al. (2006) were less conclusive. Intriguingly, despite using similar doses and measurement time points, Carter et al. (2005) reported no significant effects on span length, whereas Wittmann et al. (2006) found noteworthy impairments.

**3.3.2 Effects of psilocybin on conflict monitoring IDS: 1,3,4,8**

The body of literature analyzed in this section revolves around the impact of psilocybin on conflict monitoring, primarily employing variations of the Stroop task as a measurement tool (Table 2).

The investigation by Barrett et al. (2018; ID = 1) aimed to discern the influence of psilocybin and dextromethorphan (DXM) on emotional conflict monitoring. Four hours after administering the compound, participants undertook an emotional Stroop task. Although accuracy remained unaffected, there was a noticeable, dose-dependent increase in reaction time within the psilocybin condition. The longest response time was observed at the highest dose (∆ = 59.7ms, p < 0.0001). The authors postulate a speed-accuracy trade-off as a possible explanation for these findings (Barrett et al., 2018).

Next, Cavanna et al. (2022; ID = 3) conducted a semi-naturalistic study exploring the effects of low (micro) doses of Psilocybe cubensis on a range of factors including behavior, creativity, perception, and cognition. Employing a double-blind, placebo-controlled design, the study involved 34 participants and utilized ground edible mushrooms as a placebo. Cognitive assessments were made 180 minutes after capsule administration. While no significant differences in Stroop task performance were found between the psilocybin and placebo groups, reaction time did increase in the incongruent condition, reaching statistical significance at p < 0.005, albeit only before correction for multiple comparisons (Cavanna et al., 2022).

In a specialised context, Doss et al. (2021; ID = 4) explored the long-term effects of psilocybin treatment on conflict monitoring in patients diagnosed with Major Depressive Disorder (MDD). This open-label clinical trial involved 24 participants, administering either a medium dose of 20 mg/70 kg or a high dose of 30 mg/70 kg of psilocybin. Assessments were made at multiple time points: 8 weeks before, at baseline, and 1 and 4 weeks after the treatment. The Stroop task results indicated no significant impact of psilocybin on either reaction time or accuracy.

Lastly, the research led by Quednow et al. (2011; ID = 8) examined the cognitive ramifications of psilocybin, using a double-blind, placebo-controlled, crossed, counterbalanced design. In this study, 16 participants were administered 18 mg/70 kg of psilocybin, and their cognitive performance was assessed 85 minutes later using the Stroop task. The study revealed an increase in both error rate and reaction time in the conflict condition, effects which were reversible upon administration of Ketanserin, a 5-HT2A antagonist (Quednow et al., 2011).

In summary, the extant literature offers a nuanced understanding of psilocybin's effects on conflict monitoring, as measured by the Stroop task. Barrett et al. (2018) identified a dose-dependent increase in reaction time without affecting accuracy. Cavanna et al. (2022) found an increase in reaction time, but their results did not withstand correction for multiple comparisons. On the other hand, Doss et al. (2021) observed no significant effects in patients with MDD. Quednow et al. (2011) noted an increase in both error rate and reaction time, which was mitigated by Ketanserin.

**3.3.3 Effects of psilocybin on inhibition IDs: 3,7, 11**

The following discussion reviews three key studies that employed the Go/NoGo task (Table 2) as a framework to assess inhibitory control under the influence of psilocybin.

Kometer et al. (2012; ID = 7) carried out a double-blind, within-subject, placebo-controlled randomized study involving 17 participants. They were administered either a low dose of 15.05 mg/70 kg of psilocybin, a placebo, or 50 mg of ketanserin. Measurements were obtained using the emotional Go/NoGo task 120 minutes post-drug administration. Concurrently, the study also recorded Go/NoGo specific event-related potentials (ERPs), specifically the N2 and P300, using EEG. The results indicated that psilocybin significantly extended reaction times for correct responses, an effect further modulated by the valence of the words used in the task. In particular, the psilocybin group exhibited longer reaction times for negative words compared to positive ones. This suggests a bias towards positive emotional processing under the influence of psilocybin. Additionally, error rates escalated significantly for neutral words, irrespective of the Go/NoGo condition, and this increase was not mitigated by ketanserin. Notably, psilocybin also dampened the amplitudes of both the N2 and P300 component. The authors concluded that these findings point to psilocybin's role in promoting positive emotional processing across multiple psychological metrics, and that the effects could be counteracted by ketanserin, implicating a specific role for 5-HT2A receptors (Kometer et al., 2012).

Cavanna et al. (2022; ID = 3) incorporated the Go/NoGo task in their microdosing study—design and methodology of which have been previously discussed—and found that psilocybin had no discernible impact on either reaction time or accuracy. This finding is consistent with another microdosing study by Marschall et al. (2021; ID = 11), which also reported no significant effect of microdosing on any of the measured outcome variables. Employing a double-blind, placebo-controlled, within-subject crossover design, this study included 75 participants and concluded that psilocybin at sub-perceptive levels does not influence inhibitory control.

In summary, standard doses of psilocybin appear to influence both reaction times and error rates in tasks designed to assess inhibitory control, such as the Go/NoGo task. However, when administered at microdose levels, psilocybin does not seem to have a discernible effect on the underlying inhibitory processes.

**3.3.4 Effects of psilocybin on cognitive flexibility/set shifting: Multiple EF? ID: 1,4,12**

Cognitive flexibility, also known as set-shifting, is an essential component of executive function. It allows individuals to adapt their cognitive strategies in response to new and unpredictable environmental conditions (SOURCE). Specifically, cognitive flexibility involves the ability to shift attention between different concepts or to manage multiple concepts simultaneously. This review concentrates on four studies that explore the influence of psilocybin on this critical cognitive function.

Barrett et al. (2018; ID = 1) utilized the Digit Symbol Substitution Task (DSST; see Table 2) to evaluate cognitive flexibility. Performance on the DSST was measured at baseline and at two-hour intervals up to six hours post-drug intake. The researchers discovered that psilocybin led to a dose-dependent decrease in the number of attempted responses at the two-hour mark. Notably, psilocybin did not affect accuracy, suggesting a trade-off between speed and accuracy (Barrett et al., 2018).

Mallaroni et al. (2023; ID = 12) similarly employed the DSST in their study, adopting a double-blind, placebo-controlled, crossover design with 22 psilocybin-experienced participants. DSST performance was assessed during the subjective peak experience, occurring between two and four hours post-intake. The study also reported a reduction in both the number of correct responses and attempts, yet no significant impact on overall accuracy was observed (Mallaroni et al., 2023).

Doss et al. (2021; ID = 4), in an open-label study previously discussed, evaluated cognitive flexibility using the Penn Conditional Exclusion Test (PCET). The findings showed a reduction in perseverative errors from baseline to one week, a change that persisted at a four-week follow-up. This study also probed the neural underpinnings of these cognitive shifts by investigating functional connectivity between the anterior and posterior cingulate cortex (ACC and PCC). While static functional connectivity (sFC) remained stable post-psilocybin therapy, dynamic functional connectivity (dFC) between the ACC and PCC significantly increased one week after psilocybin administration. Further analysis revealed a moderate correlation between this increase in dFC and the reduction in PCET perseverative errors at the one-week follow-up, thereby suggesting a complex interplay between neural and cognitive flexibility (Doss et al., 2021).

In summary, the reviewed studies offer consistent evidence that psilocybin impacts cognitive flexibility, although in subtly different ways. Both Barrett et al. (2018; ID = 1) and Mallaroni et al. (2023; ID = 12) utilized the DSST and found that psilocybin decreased the number of attempted responses without affecting accuracy. In contrast, Doss et al. (2021; ID = 4) employed the PCET and demonstrated that psilocybin led to a reduction in perseverative errors that was sustained over a four-week period. Collectively, these results underline the complex and nuanced effects of psilocybin on cognitive and neural flexibility, signaling the need for more comprehensive research.

**3.3.5 Effects of psilocybin on Attention** **IDs: 2,3,5,6,9**

Attention is a critical cognitive process that enables individuals to focus selectively on specific information while filtering out other stimuli. Given its importance in human cognition, attention has been extensively studied. This review highlights five studies that investigate the effects of psilocybin on attention.

Carter et al. (2005; ID = 2) used the Multiple-object Tracking task (refer to Table 2) to assess the impact of psilocybin on attention. The study found that the psilocybin group experienced significant difficulties in attentional tracking compared to the placebo group. Specifically, they tracked fewer dots and exhibited lower d’ scores 120 minutes after drug administration (Carter et al., 2005).

Similarly, Cavanna et al. (2022; ID = 3) employed the Attentional Blink and Trail Making Test tasks to measure attention. Although they observed a decrease in the visibility of T2 with a 300 ms lag in the Attentional Blink task after psilocybin administration, this result became non-significant upon correction for multiple comparisons. Moreover, the study did not find any significant differences between groups in the outcome measures of the Trail Making Test (Cavanna et al., 2022).

In another double-blind, placebo-controlled experiment, Gouzoulis-Mayfrank et al. (2002; ID = 5) used the Covert Orienting of Attention Task (COVAT) to evaluate attention. Conducted 75-90 minutes post-drug intake, the study revealed that the psilocybin group had significantly longer reaction times when compared to the placebo group, suggesting a decline in attention (Gouzoulis-Mayfrank et al., 2002).

Hasler et al. (2004; ID = 6) conducted a double-blind, within-subject dose-effect study with 9 participants and four different dosages. The Frankfurt Attention Inventory (FAIR) was employed to investigate attentional performance. The study found that psilocybin administration resulted in a significant decline in P and C scores at both medium and high doses, highlighting an impairment in attention (Hasler et al., 2004).

Vollenweider et al. (2007; ID = 9) also utilized the Frankfurt Attention Inventory (FAIR) task in a double-blind, placebo-controlled, counterbalanced study. They found that psilocybin significantly impaired attentional performance, evident from reduced P and C scores. The impairment was significant across low, medium, and high doses of psilocybin and persisted during both the peak and post-peak effects (Vollenweider et al., 2007).

In summary, existing research consistently demonstrates that psilocybin impairs various aspects of attention. Across a range of tasks and methodologies, the evidence robustly supports the notion that psilocybin administration leads to significant impairments in attentional performance.

**3.3.6 Conclusion of review**

In conclusion, the effects of psilocybin on executive functions such as inhibitory control, cognitive flexibility, and attention are complex and multifaceted. Research on inhibitory control has shown that while standard doses of psilocybin alter reaction times and error rates, microdosing seems to have no discernible impact on these measures (Cavanna et al., 2022; Kometer et al., 2012; Marschall et al., 2021). In the realm of cognitive flexibility, psilocybin affects performance on tasks like the DSST and PCET, without significantly altering accuracy, and is associated with lasting reductions in perseverative errors and changes in neural connectivity (Barrett et al., 2018; Mallaroni et al., 2023; Doss et al., 2021). Finally, with respect to attention, there is a consistent indication of impairment across various tasks and doses, as evidenced by multiple studies (Carter et al., 2005; Cavanna et al., 2022; Gouzoulis-Mayfrank et al., 2002; Hasler et al., 2004; Vollenweider et al., 2007). Taken together, these studies indicate thatpsilocybin impairs cognitive processes at higher doses. Further research is needed to deepen our understanding of the dose-response relationships and the neural mechanisms underpinning these effects, thereby informing both clinical applications and theoretical models of cognitive neuroscience.

**3.4 Results of the meta-analysis**

The present meta-analysis assessed the effects of psilocybin on cognition across different dosages and cognitive categories. We employed both random-effects and multivariate meta-analysis models to analyze in total 34 effect sizes from 10 unique studies, in which reaction time (RT) or Accuracy (ACC) was measured as a proxy for cognitive performance.

Multiverse analysis:

* Decision points
  + Peak timing
  + MD inc vs exc

Decision tree figure

Moderation

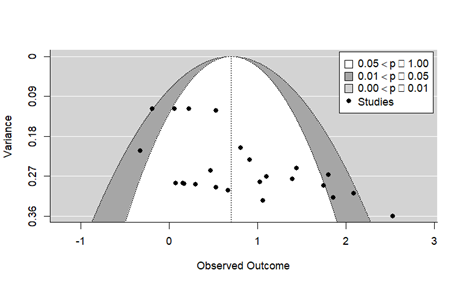
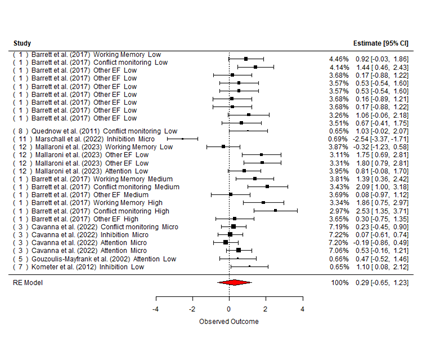
| Peak | | Microdosing | |
| --- | --- | --- | --- |
| Included | Excluded |
| 2-3h | |  | 0.8973 (0.6975-1.0972) |
| 2-4h | |  |  |
| 1-4h | |  |  |

**3.4.1 Reaction time**

**3.4.1.1 Nested & non nested model**

The dataset of reaction times included 27 Effect sizes from 7 unique studies. We decided to use the measurement timepoint of 180mins as a threshold in the moderation analysis, to analyze the effect of measurement time point on RT. This resulted in 17 effect sizes where the measurement was conducted prior to 180 minutes post-administration and 10 effect sizes post-180 minutes.

A multilevel random effects meta-analysis, accounting for the nesting of multiple effect sizes within the same study, revealed a non-significant medium overall effect size (ES) of Hedges’ g = 0.29 (SE = 0.48, z = 0.6, p = 0.55, 95% CI [-0.65, 1.23]). The heterogeneity was substantial (I2 = 86.45%) and significant (Q(26) = 110.52, p < .001). One outlier was excluded based on the standardized residuals, with a threshold of |2| (ID = 11). The result of the analysis after the exclusion of the outlier showed a significant overall ES of Hedges’ g = 0.7 (SE = 0.18, z = 3.86, p < 0.001). The heterogeneity was reduced (I2 = 30.38%), but still significant (Q(25) = 55.95, p < .001).



**Figure 1** RT left panel funnel plot, right panel forest plot

For robustness we also evaluated the nonnested random-effects model, which revealed a significant overall effect size (ES) of Hedges’ g = 0.76 (SE = 0.14, z = 5.31, p < .001), with a 95% confidence interval ranging from 0.48 to 1.04. The estimated heterogeneity was substantial (I2 = 55.32%, Q(25) = 55.95, p < .001). This pooled effect size is similar to the multilevel pooled effect size, suggesting a robust interpretation of the magnitude of the effect. The formal test might however be biased towards significance (too small standard errors) due to ignored nesting effects.

**3.4.1.2 Publication Bias**

For the model including the outlier, the rank correlation test for funnel plot asymmetry indicated some evidence of asymmetry suggesting of publication bias (Kendall's τ = 0.41, p = .003). The fail-safe N calculations using the Rosenthal, Orwins and Rosenberg approaches indicated that a substantial number of studies with an average sample size and null result would be required to negate the observed effects (FSNRosenthal = 401, FSNRosenberg = 204). These fail safe numbers exceed the threshold of 5\*k + 10 = 180 studies, indicating evidence against publication bias. According to Orwin’s fail safe number, FSNOrwin = 160 studies of average sample size and null results would be required to reduce the current pooled effect to a trivial Cohen’s d level of 0.1.

For the model excluding the outlier, the rank correlation test for funnel plot asymmetry indicated no evidence of asymmetry suggesting absence of publication bias (Kendall's τ = 0.41, p = .003). The fail-safe N calculations using the Rosenthal and Rosenberg approaches indicated that a substantial number of studies with an average sample size and null result would be required to negate the observed effects (FSNRosenthal = 565, FSNRosenberg = 324). These fail safe numbers largely exceed the threshold of 5\*k + 10 = 105 studies, indicating evidence against publication bias. According to Orwin’s fail safe number, FSNOrwin = 186 studies of average sample size and null results would be required to reduce the current pooled effect to a trivial Cohen’s d level of 0.1.

**3.4.1.3 Moderation for RT**

**3.4.1.3.1 Dosage category as moderator**

Since significant heterogeneity was detected, the effect of psilocybin on RT might differ between dosages (low, medium, high dosage). To this end, moderation analyses were conducted to examine the impact of dosage on the effect sizes. In the first analysis, dosage did emerge as a significant moderator (QM(2) = 6.66, p = .034) with significant heterogeneity of residuals (Qresidual(16) = 45.64, p = .003) . The effect size at low dosage (reference category) was significant, B = 0.62, SE = 0.17, z = 3.67, p < .001, 95% CI [0.29, 0.95]. Comparison of the mid dosage to the low dosage did not reveal a significant difference in effect sizes, B = 0.53, SE = 0.35, z = 1.52, p = .13, 95% CI [-0.15, 1.21]. However, the high dosage showed a significantly different effect size compared to the low dosage, B = 0.85, SE = 0.36, z = 2.34, p = .02, 95% CI [0.14, 1.56], suggesting a substantial increase in effect size from low to high dosage.

In a subsequent analysis, the dosage categories were combined into two levels: low, and mid & high combined. Also this model showed an overall significant moderation with heterogeneity of residuals (QM(1) = 6.16, p = .013; Qresidual(24) = 46.13, p = .004 ). The effect size for the low dosage remained significant, B = 0.62, SE = 0.17, z = 3.67, p < .001, 95% CI [0.29, 0.95]. The combined mid and high dosage category demonstrated a significant difference in effect size compared to the low dosage, B = 0.68, SE = 0.27, z = 2.48, p = .013, 95% CI [0.14, 1.22]. This indicates that increasing the dosage from low to a combined mid and high level is associated with a significant increase in the effect size.

**3.4.1.3.1 Pre vs post peak as moderator**

Pre or post peak timing (180mins) as moderator yielded no significant moderation (QM(1) = 0.23, p = .63).

**3.4.1.3.2 Cognitive function as moderator**

In our moderation analysis focusing on the impact of cognitive functioning categories, we observed significant residual heterogeneity (QE = 46.71 with 21 degrees of freedom, p = 0.001), suggesting substantial variability among studies. The overall test of cognitive functioning as a moderator approached significance (QM(4) = 9.27, p = 0.055). Specifically, the intercept, or the effect size for the reference category of cognitive functioning, was found to be significant (B = 0.68, SE = 0.30, z = 2.24, p = 0.025). For the category of 'Conflict Monitoring', the effect size showed a trend towards significance (B = 0.56, SE = 0.33, z = 1.72, p = 0.085). The categories of 'Inhibition' (B = 0.07, SE = 0.37, z = 0.18, p = 0.855), 'Other Executive Function' (B = -0.34, SE = 0.35, z = -0.96, p = 0.335), and 'Working Memory' (B = -0.15, SE = 0.40, z = -0.39, p = 0.696) did not demonstrate significant differences in effect sizes compared to the reference category.

**3.4.2 Accuracy**

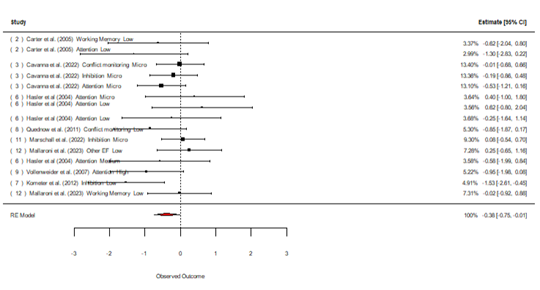
**3.4.2.1 Nested & non nested model**

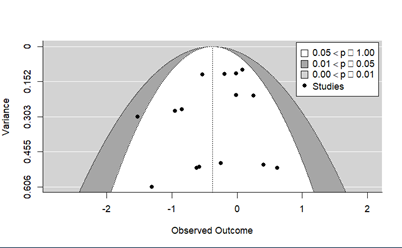
A multilevel meta analysis on the subset of 15 accuracy effect sizes from 8 unique studies revealed a significant negative overall pooled effect size of Hedges’ g = -0.38 (SE = 0.19, z = -2, p = .0.045 95% CI [-0.75, -0.009]). No heterogeneity was detected (I2 = 35.6%, Q(14) = 16.66, p = .275)

For robustness we also evaluated the non-nested random-effects model, which again revealed a overall effect size (ES) of Hedges’ g = -0.3 (SE = 0.138, z = -2.165, p = .003), with a 95% confidence interval ranging from -0.57 to -0.028. The estimated heterogeneity was small (I2 = 15.97%, Q(14) = 16.66, p < .27). This pooled effect size is similar to the multilevel pooled effect size, suggesting a robust interpretation of the magnitude of the effect. The formal test might however be biased towards significance (too small standard errors) due to ignored nesting effects. No outlier based on the standardized residuals, with a threshold of |2|.

**3.4.2.2 Publication Bias**

For the model including the outlier, the rank correlation test for funnel plot asymmetry indicated some asymmetry, but was not significant (Kendall's τ = -0.26, p = .2). However, the Fail-safe N calculations yield mixed results. Using the Rosenthal approach, the Fail-safe N was 21, suggesting that 21 additional studies with null results would be required to bring the observed significance level (p = .0057) above the conventional alpha of .05. This number is below the threshold of 85 (calculated as 5k+ 10), indicating a potential vulnerability to publication bias. The Orwin method yielded a Fail-safe N of 40, indicating that 40 additional studies with an effect size of -0.10 would be necessary to bring the average effect size down from -0.3645 to the specified target of -0.10. This number is also below the threshold of 85, suggesting a similar vulnerability. In contrast, the Rosenberg approach indicated a lower level of robustness, with a Fail-safe N of only 6, far below the recommended threshold. This implies that adding six null-effect studies could increase the observed significance level (p = .0222) above the alpha threshold of .05. These mixed findings imply that while there isn't strong evidence of publication bias according to kendal’s tau, the possibility cannot be entirely dismissed, especially considering that all fail-safe N values are below the calculated threshold of 85, suggesting a potential susceptibility to publication bias. Thus, in addition to the Fail-safe N calculations, a modified Egger's test was conducted to further assess the potential for publication bias. The Egger's test, which uses precision as a moderator in a multivariate meta-analysis framework, did not reveal significant evidence of publication bias (QM = 1.4361, df = 1, p = 0.2308). This suggests that the effect sizes in the analysis are not disproportionately influenced by study size or precision, a common indicator of publication bias.

****

****

**Figure2:** ACC. Left: Forest plot Accuracy. Right, Funnel plot Accuracy

**3.4.2.3 Moderation for Accuracy**

**3.4.2.3.1 Dosage category as moderator**

An analyses was conducted to examine the impact of dosage on the effect sizes of accuracy. In the first analysis, dosage did not emerge as a significant moderator (QM(2) = 1.27, p = .53) with non-significant heterogeneity of residuals (Qresidual(12) = 14.74, p = .26). In the subsequent analysis, the dosage categories were combined into two levels: low, and mid & high combined. Also this model showed an overall non-significant moderation with non-significant heterogeneity of residuals (QM(1) = 1.3, p = .25; Qresidual(13) = 14.91, p = .31 ).

**3.4.2.3.2 Pre vs post peak as moderator**

Pre or post peak timing (180mins) as moderator yielded no significant moderation (QM(1) = 0.05, p = .82).

**3.4.2.3.2 Cognitive function as moderator**

The analysis with cognitive function category as moderator was also non-significant (QM(4) = 1.69, p =.79), with no residual heterogeneity (QE(10) = 14.65, p =.15).

**3.5 Results summary**

For RT, our multilevel random effects meta-analysis initially indicated a non-significant medium overall effect size (Hedges’ g = 0.29). However, after excluding one outlier, the effect became significant (Hedges’ g = 0.7), with reduced heterogeneity. The nonnested model supported these findings, showing a significant overall effect size (Hedges’ g = 0.76). Notably, substantial heterogeneity was observed, and the publication bias assessment revealed mixed results. While Kendall's τ suggested asymmetry in the original model, the fail-safe N calculations for both models (FSNRosenthal = 401/565, FSNRosenberg = 204/324, FSNOrwin = 160/186) exceeded the threshold of 5\*k + 10 (85/105), arguing against publication bias.

For ACC, the multilevel meta-analysis revealed a significant negative overall pooled effect size (Hedges’ g = -0.38), which was corroborated by the nonnested model (Hedges’ g = -0.3). Heterogeneity was small to moderate and not significant. The publication bias assessment indicated no significant asymmetry and mixed results in fail-safe N calculations (FSNRosenthal = 21, FSNRosenberg = 6, FSNOrwin = 40), with all values below the threshold of 85, suggesting a potential risk of bias.

These findings illustrate that psilocybin may have a more pronounced and variable impact on reaction time than on accuracy. While dosage appears to be a significant moderator for RT, no such effect was observed for ACC. The presence of heterogeneity and the mixed results regarding publication bias, particularly for RT, necessitate cautious interpretation. The exploratory nature of this investigation highlights the need for further research, especially with a more diverse range of cognitive measures and controlled dosing conditions, to better understand psilocybin's cognitive effects.

**4. Discussion**

* is the selection of cognitive tests currently the best way to assess cognition
* Aufruf: kognitive research with psychedelics, wie sollte man das ändern? differenz nuance
* acute effects not, but long term effects not yet shown
* on repetitive regimes
* is cognitive testing under tripping rather a dual task? Dual task hypothesis or do individuals work under harder conditions, but better?

**4.x Limitation**

* only acute effects, not long term effects

**5. Further information**

**5.1 Registration and protocol**

For the present study, no prior hypotheses were set. Furthermore, the present study was not preregistered, and also no study protocol was produced prior to data collection and analysis, thus, the present study is considered exploratory, and should not be used to generalize.

**5.2 Author contributions**

The conceptualization of the research was collaboratively developed by SEG, MVE, and ML. The methodology was designed by PY, RR, and SEG, with PY and RR also responsible for the software development and validation. Formal analysis was conducted by PY and RR, with PY leading the investigation. Resources were provided by SEG and MVE, and data curation was handled by PY and FO. The original draft of the manuscript was written by PY, and the manuscript was reviewed and edited by PY, MVE, and ML. Visualization was executed by PY and RR. Supervision of the project was overseen by SEG and MVE. Project administration was conducted by PY, SEG, and MVE. Funding acquisition was solely managed by SEG.

**5.3 Availability of data and code**

All the data and code underlying the present study are available [here](https://drive.google.com/file/d/1gvfnyh_3qC3pwXFPp0aD7vQE4a45dUEM/view?usp=sharing) as supplementary material. We adhere to OPEN guidelines and encourage the usage of our data for further exploration.

**5.4 Conflict of interest**

The authors declare no conflict of interest.