A Replication of the White Bear Experiment: Investigating the Link Between Depression Scores and Mood Rebound in Suppressed Thought Expression

Niels Harkema

s4351703

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Supervisor: Dr. Ineke Wessel

Second evaluator: Dr. Norbert Borger

Group number: 20

In collaboration with: Bert van der Leest, Bo Kahmann, Mailer Al-Rabadi, Evelina Stefanova

and Yva Scheibner.

February 17, 2025

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned

(keep the above text as it is)

Abstract

The "White Bear" effect refers to the paradoxical nature of the resurgence of suppressed thoughts. Furthermore, a connection is created between the suppressed thought and the associated mood, whereby the activation of the thought leads to the reinstatement of the associated mood (Wenzlaff et al., 1987,1991). Although, replication studies of these claims have shown inconsistent results. To examine whether the original findings from Wenzlaff et al., (1991) is replicable, a group study was conducted replicating the second experiment. In the experimental setting, 61 non-clinical, first-year students were studied in regard to thought suppression. The results find no support for the "White Bear" effect or the mood reinstatement. However, the results were impacted by methodological limitations. On the other hand, support was found for the moderating relation between depressive severity and thought suppression on mood change. Future research may examine the role of thought suppression as a maladaptive coping strategy for individuals with high scores on depressive severity.

Keywords:

Depression, Mood Reinstatement, Replication, Thought suppression, White Bear effect.

A Replication of the White Bear Experiment: Investigating the Link Between Depression Scores and Mood Rebound in Suppressed Thought Expression.

Carl Jung (2014, p.27) once said "*We also know that it is dangerous to suppress it, because the unconscious is life and this life turns against us if suppressed*", which might be especially relevant in relation to mental health and thought suppression. According to the World Health Organization (WHO., 2022) approximately 970 million people worldwide experienced mental health disorders, primarily anxiety and depression disorders in 2019. Furthermore, the WHO reports that the COVID-19 pandemic led to an estimated 28% rise in major depressive disorders within one year. Dejonckheere (2017) suggests that a potential reason for the stigma surrounding mental health and the limited access to depression care, despite the availability of effective treatments, is the excessive focus on the pursuit of happiness in Western society. The increasing rates of depression have been associated with Western societal overemphasis on happiness. The perceived societal denouncement of negative emotions is a driving force behind instigating depression and predicts an increase in overall severity and most individual symptoms of depression, which possibly links to the idea that the suppression of negative thoughts play a role in the worsening of depression.

White Bear Experiment

Psychological research into thought suppression is greatly influenced by the "White Bear" studies (Wegner et al., 1987), which introduced evidence for the paradoxical nature of thought suppression. In the experiment, where participants were instructed to either suppress or express it was found that the suppression group reports a higher amount of target item thoughts. The paradoxical finding that the participants who are asked to suppress their thoughts about the target items have an increase in target item thoughts at a later moment compared to people who are freely allowed to express their thoughts, is called the "White Bear" effect. Further research (Wenzlaff et al., 1991) found that thought suppression not only creates a bond with the suppressed thought, but also with the associated mood state, whereby the activation of the thought leads to the reinstatement of the mood. The study found this connection with two experiments, where participants had their mood induced by music, while also being asked to either suppress or express their thoughts about a white bear. The results of the first experiment found that when participants were asked to think about a white bear, those in similar moods during thought suppression and expression showed a greater rebound of the target item. The results of the second experiment found that participants who had initially tried to suppress their thoughts experienced a reinstatement of the same mood when asked to express their target item at a later moment.

Replication

A variety of studies have replicated the "White Bear" effect (Clark et al 1991, 1993; Wenzlaff et al., 1991). However, other studies have found only partial support (Kelly & Kahn 1994; Rutledge et al., 1993, 1996), while others have even failed to find the "White Bear" effect at all (Merckelbach et al., 1991; Roemer & Borkovec 1994).

Some studies suggest that the possible reason that they have not replicated the "White Bear" effect has to do with individual differences influencing the effect (Rutledge et al., 1993, 1996; Smari et al., 1994). One of the individual differences is the connection between thought suppression and various psychopathologies, particularly generalized anxiety disorder and depression (Purdon., 1999).

Individual difference: Depression

Upon further exploring the individual difference of depression in the literature, it shows that individuals with a recent depressive episode suppress their depressive thoughts after their recovery, seemingly to preserve their emotional well being (Wenzlaff et al., 1998). This mechanism is shown in an experiment (Wenzlaff et al., 1998) involving unscrambling sentences that could form either positive or negative statements. The introduction of thought suppression to the experiment led to an increase in the number of negative statements by participants recovering from a recent depressive episode. They formed negative statements at approximately the same rate as participants with an active depression diagnosis. The experiment suggests that depressed individuals have considerable difficulty inhibiting their negative thoughts, despite the fact that they often expend considerable time and effort trying to do just that (Wenzlaff et al., 1993; Wenzlaff & Bates 1998). An explanation is that individuals with depression have more access to distractors that are associated with the negative mood state, which they are trying to suppress. These findings tie in with the mood reinstatement by Wenzlaff et al (1991), that mood-related distractors create an association between the suppression target and the relevant mood.

A recent study (Silva, 2018) explored how individuals with depressive symptoms attempt to suppress worrisome thoughts and the impact on mental health. The study involved participants with sub-clinical and non-clinical depressive symptomatology, participating in thought suppression of worrying thoughts. The results found that in the context of depressive symptoms, the effort to suppress worrisome thoughts may make those thoughts more persistent. The study (Silva, 2018) supports the "White Bear" effect, as it shows the same effect regarding mood reinstatement. Furthermore, it found that the suppression of worrisome thoughts may exacerbate depressive symptoms.

Further research gives a possible reasoning behind the relation between depression and thought suppression. The study by Rosebrock et al., (2019) investigated how rumination interacts with thought suppression in a clinical population. The clinical population consisted of veterans with unipolar depression, both with and without comorbid PTSD. The results suggest that while rumination is present in both groups, its function may differ. In depression alone, rumination typically centers around self-criticism and failure-related thoughts, whereas in comorbid PTSD, it often involves trauma-related content. Importantly, the suppression of negative thoughts paradoxically makes them more accessible, feeding into the cycle of ruminative thinking that sustains depressive symptoms. This supports Wenzlaff et al., (1993, 1998) earlier claims about the difficulty of inhibiting negative thoughts for individuals with depression. This study highlights that rumination is a central cognitive process in depression, and that targeting cognitive avoidance strategies like thought suppression may help reduce rumination and therefore alleviate depressive symptoms. Thus thought suppression plays a crucial role in exacerbating the cognitive patterns that increase depression.

In summary, the foundation laid by the work of Wegner and Wenzlaff (1987,1991) inspired a myriad of new studies, further exploring the paradoxical effect of thought suppression. As some studies have found support(Clark et al 1991, 1993), others have found partial (Kelly & Kahn 1994; Rutledge et al., 1993, 1996) or no support for the "White Bear" effect (Merckelbach et al., 1991; Roemer & Borkovec 1994). However, the prevailing literature shows that suppression leads to the "White Bear" effect and the mood reinstatement. Three key claims emerge from this literature. Firstly, the suppression of thoughts results in a rebound effect, where those thoughts resurface with greater intensity (Wegner et al., 1987;Wenzlaff et al., 1991). Secondly, suppressed thoughts are associated with the mood in which they were suppressed. This means that the resurgence of the suppressed thought will lead to the reinstatement of the original mood state (Wenzlaff et al., 1991; Purdon, 1999; Silva, 2018). Lastly, thought suppression shows a stronger negative effect on mood in individuals with depression (Wenzlaff et al., 1993, 1998; Silva, 2018; Rosebrock et al., 2019).

Research Question

While expanding further research is crucial for broadening established claims, it is equally important to critically examine those established claims. As reflected by the

discrepancies in replications of the white bear effect (Kelly & Kahn 1994; Rutledge et al 1993, 1996; Merckelbach et al., 1991; Roemer & Borkovec 1994), expanding on claims and replicating existing literature is important to allow for the inferences about reliability and generalizability of findings. The open science collaboration (2015) found that in replicating 98 previously published studies that only 40% produced the same result. This emphasizes the importance of replication to ensure that psychological claims are robust and can be built and expanded on further. Furthermore, a study (Shrout & Rodgers, 2018) reflecting on replication supports the idea that replication is crucial for refining theories and establishing trustworthy psychological knowledge, especially when discrepancies exist.

Regarding the discrepancies in replicating the effects found in the literature, this study aims to replicate the claims of the "White Bear" experiment and mood reinstatement based on the second experiment by Wenzlaff et al (1991). The focus on the second experiment allows for the assessment of the relationship between thought suppression, mood, and mood reinstatement. Allowing the experiment to replicate the "White Bear" effect and the mood reinstatement as earlier explained in the literature (Wegner et al., 1987; Wenzlaff et al., 1991). Additionally, this paper explores the relationship of thought suppression, depressive severity and mood change to further investigate whether individuals with depression show a stronger relation between thought suppression and mood. The study aims to explore this claim in a non-clinical population by examining the relation between depressive severity and thought suppression on mood.

Research Question 1: Does the mood, experienced during the thought-suppression session, become reinstated when the previously suppressed thought is reinstated?

"White Bear" Effect

1. Participants exhibit fewer target item thoughts in the suppression condition compared to those in the expression condition.

2. Participants report more thoughts of the target item in the second phase of the experiment if they were assigned to the suppression condition, as compared to the expression condition.

Mood Reinstatement

3. Participants exhibit lower mood ratings in the negative mood induction condition compared to those in the neutral mood induction condition.

4. Participants exhibit lower mood ratings in the negative suppression condition compared to the two neutral conditions and the negative expression condition in the second phase of the experiment.

Research Question 2: How does the severity of depression influence the effect of thought suppression on mood change?

Individual Difference: Depression

5. Participants in the suppression conditions show a stronger negative relation between QIDS-SR scores and the positive mood change scores in the suppression condition when compared to the positive mood change scores of the expression condition.

6. Participants in the suppression conditions show a stronger positive relation between QIDS-SR scores and the negative mood change scores in the suppression condition when compared to the negative mood change scores of the expression condition.

Method

Participants

Eligible participants were 61 (23% Male, 77% Female, 0% nonbinary/other) first-year students in Psychology at the University of Groningen, who participated in return for a credit in a psychology course *(A Practical Introduction to Research Methods PSBE1-28)*. Their mean age was 20.17 (range 18-31, SD=2.51). The highest percentage of participants were from the Netherlands (46%) and Germany (13%).

The study applied two exclusion criteria, one before the experiment where participants were asked if they were diagnosed or in treatment for any mental disorder. In the case that a participant was in treatment or diagnosed, they would be excluded. Secondly, participants were excluded if they failed to meet one or both of the attention validity checks included in the questionnaires. Initially, N=64 participants were screened, three participants were excluded from the sample for not meeting the attention validity check in the questionnaire.

The study received approval from the Ethics Committee of the Faculty of Behavioral and Social Sciences at the University of Groningen. The research code is PSY-2425-S-0047.

Research Design

The study employs a 2 by 2 design, with two phases, whereby both conditions are employed in the first phase. The participants were assigned to a thought instruction condition (Suppression, n=27 or Expression, n=34) and a mood induction condition (Negative, n=31 or Neutral, n=30). This led to the participants being randomly assigned to one of four conditions in the first phase, leading to Suppression/Negative (n=9), Suppression/Neutral (n=18), Expression/Negative (n=22) and Expression/Neutral (n=12). The second phase was the same for all conditions, whereby all participants were asked to express their thought of the target item.

The power analysis showed, using G*Power (Faul et al., 1996), that a total of 119 participants were required to reach the desired power of .90 to detect a true effect, to reject the Null hypothesis with alpha of .05, while aiming for Cohen's f = 0.3.

Materials

Positive and Negative Affect Measure

The study made use of a modified version of the Positive and Negative Affect schedule (PANAS; Watson et al., 1988) to measure the positive and negative mood rating through self report. This modified version of the scale includes the original 10 positive-affect items (e.g. determined, excited, interested) and 10 negative-affect items (e.g. irritable, ashamed, afraid). The scale also included three items assessing neutral mood, as well as three items which were used in the original study (Wenzlaff et al., 1991), which amounts to a total of 26 items assessing mood. Lastly, it contains three items used as distractor items from the original study.

All items were rated on a 5-point-Likert-scale ranging from (1) "very slightly or not at all" to (5) "extremely". The Adjusted PANAS reliability scores were assessed for the 1-10 positive affect items (Alpha = .61) and 11-20 negative affect items (Alpha = .75).

Reported Thought Measurement

Self report is used to measure the occurrence of "White Whale" thoughts, whereby each participant was given one A4 paper where they were instructed to write one check mark for each "White Whale" thought. The front of the paper indicated "*phase 1*" and was used to tally the marks in the first phase, while the back of the paper indicated "*phase 2*" and was used to tally the marks in the second phase. Once the data collection was completed, the tally marks were individually counted by different researchers to verify the number of reported "White Whale" thoughts. These were manually entered into the collective data file, allowing for further analysis.

Mood Induction

The first phase of the thought report included the mood induction, in which either negative inducing music (Russia Under the Mongolian Yoke, Field of the Dead, from Alexander Nevsky, Op. 78, by Prokofiev) or neutral inducing music (Common Tones in Simple Time, by John Adams) would be played over the headphones of the participant. The music choice for the negative and neutral mood induction was adopted from the original study (Wenzlaff et al., 1991).

QIDS-SR

The severity of depressive symptoms was measured through the use of a self report in Qualtrics. The self report used was the Quick Inventory of Depressive Symptomatology and Self-Report (QIDS-SR)(Rush et al., 2003), consisting of a 16-item self-report questionnaire used to assess the severity of depressive symptoms. The questions were recorded on a likert-scale ranging from 0 until 3. It covers key symptoms of major depressive disorder in line with the DSM-IV. The QIDS-SR total scores range from 0-27, whereby the total score is obtained by adding the highest scores of each symptom domain. A higher score indicates a higher level of depressive severity, the QIDS-SR demonstrated a high internal consistency (Cronbach's Alpha = .793).

Other instrument

As this study was part of a group study, other instruments were used, but not further explored in this individual study. Firstly there were two tasks, Go/No-Go and OSPAN, and the questionnaires of the WBSI, Padau and RRS. Qualtrics is the computer platform used in the experiment, with which the following was measured; the questionnaire responses, the PANAS scores and descriptive information of the participants (Qualtrics, 2025).

Procedure

The study was conducted in English by six students from the Bachelor Psychology in

Groningen. Collection of data ran from 28 November 2024 until 20 December 2024. The experimenter training included creation of a script, acting out the experiment, reflecting upon possible improvements and multiple mock trials with practice participants. All computer programs and papers needed for the experiment were inspected by all research members.

The participants signed up for participation through the University recruitment platform SONA. The advertisement for the study framed the study as "Binding Thought and Music". The individuals were individually tested in a laboratory cubicle, whereby participants received instructions during the experiment through the Qualtrics system. To minimize confusion around the Qualtrics instructions, the participants also received instructions from the researchers. The door to the cubicle was closed and the researcher left the cubicle during each task to give the participant more privacy and limiting distractions. The participants would start with the questionnaires, which includes the QIDS-SR. Once the participant left the cubicle to inform the researcher of the completion of the questionnaire task, the researcher would guide the participant to the dedicated Go/No-Go (Tucha et al., 2013) test cubicle. After successful completion of the trial, the researcher left the cubicle and the participant completed the task individually. When the participant completed the task, the researcher guided them back into their cubicle to perform the thought report session 1 and mood report session 1.

A passcode was required before thought report session 1 could begin, this page also included either an "s" or an "e" in the text, allowing the researcher to know which thought instruction condition they had to explain. The allocation of the thought instruction for "s" or "e" was randomized by Qualtrics. Additionally, the participants were asked to wear headphones during the 9-minute thought report session as the mood induction music would be played over the headphones. The negative or neutral mood induction was randomly assigned by Qualtrics. Following the thought report session 1, the page would continue to the mood report page, where the participant could report their mood.

The mood reporting was followed by the OSPAN task (OSPAN; Turner & Engle, 1989). The researcher explained the task and was present for the duration of the test trial to ensure that the participant understood the task. After completion the researcher would instruct the participant to start on the second thought report phase. The participant was instructed to express their thoughts freely and did not receive any mood induction. After completing the second 9-minute thought report session they reported their mood for the second time. This concluded the experiment.

Data Analysis

The data was analyzed using the software JASP (Version 19.3.0; JASP team, 2025). The first step was to transform the raw data to allow for easier analysis of the following variables: QIDS-SR total, condition groups and mood scores. The QIDS-SR was transformed from the description of the score to a likert-scale of 0-3 which is in keeping with the official scoring. In contrast to other symptoms, the symptoms of sleep, weight/appetite and psychomotor agitation were measured by multiple questions, which leads to multiple measurement points. To obtain the score for this specific symptom, the highest value was taken and used as a single representation for that symptom. The total QIDS-SR score was calculated by combining the score of each individual symptom, which led to a score between 0-27.

The second variable created was called "Conditions", to represent the interaction between the thought instruction and the mood induction, allowing for comparison between the four condition groups. This leads to the following groups within the condition variable; Suppression-Negative, Suppression-Neutral, Expression-Negative, Expression-Neutral.

The last variables created were the mood change scores for positive and negative mood, which were calculated by subtracting the mood score of phase 2 from the mood score

of phase 1 for both positive and negative scores. The positive mood scores consist of scores 1-10 on the PANAS test, while the negative mood scores consist of scores 11-20 on the PANAS test. The mood change scores for positive and negative mood were created to allow for the analysis of the mood change between phase 1 and 2 in both positive and negative affect.

The independent variables were the experimental phases, conditions (negative or neutral mood and suppression or expression instruction) and the total score on the QIDS-SR. The dependent variables are the total number of reported target thoughts in phase 1 and phase 2, the total mood score of phase 1 and phase 2 and the mood change scores.

The normality assumption has been analyzed by the use of the Shapiro-Wilk test and homogeneity assumption has been analyzed using the Levene's test. Furthermore the independence of observations was ensured by the random assignment of the participants to the conditions by the Qualtrics system as described in the Method section. The researchers were blind to the mood induction condition and were not in control of the thought instruction selection. A preliminary visual analysis of the variables, through the use of QQ-plots and Boxplots, showed that our data was not normal. Which influences the interpretation of the results.

The main analysis consists of four hypotheses that have been tested using a Welch *t*-test, except for hypothesis four which has also been tested using an ANOVA. The ANOVA allows for the direct comparison between the condition groups, while the Welch *t*-test allows for deeper understanding of specific group comparisons. The individual analysis has been tested using Pearson's correlation. The Pearson correlation allowed for the relational analysis of variables. Secondly, the correlation between suppression and expression were compared through the use of an online calculator (Lenhard, 2014) based on the calculations of Eid, M., (2011).

Results

Main Replication Results

Condition Manipulation check

Table 1 summarized the descriptives of thought report phase 1 for the instruction condition.

The Welch *t*-test tested that the reported thoughts in the expression group would have been significantly greater than those in the suppression group in the first thought reporting phase, but no significant effect was found, t(58.761)=1.346, p < .092, Cohen's d = .344. The result shows that there is no significant evidence that participants in the suppression condition exhibit significantly fewer target item thoughts compared to the expression condition as stated in hypothesis 1.

Levene's test for assumption of homogeneity of variances was met for the Welch *t*-test, F(1,59)=1.409, p=.240. The normality of the distribution of thought report phase 1 was analyzed by the Shapiro-Wilk test, which shows that the suppression condition, p=<.001, deviated significantly from normality. However, the expression condition, p=.073, does not significantly deviate from normality.

Table 1

	Suppression (n=27)	Expression (n=34)
Mean	6.37	8.088
Standard deviation	4.559	5.407
Median	5	7.5
IQR	3	5.750
Shapiro-Wilk	.769	.942
p-value of Shapiro Wilk	<.001	.073

Descriptive statistics for thought report 1 split over the instruction condition

White Bear Effect

Table 2 summarized the descriptives of thought report phase 2 for the instruction condition.

The Welch *t*-test tested that participants in the expression condition reported significantly less target item thoughts in the second phase than those in the suppression group, but no significant effect was found, t(55.251)=-.851, p < .199, Cohen's d = -.220. The results show that there is no significant evidence that participants in the suppression condition will report more target item thoughts in the second phase compared to the expression group as stated in hypothesis 2.

Levene's test for assumption of homogeneity of variances was met for the Welch *t*-test , F(1,59)=1.409, p=.240. The normality of the distribution of the thought report was analyzed by the Shapiro-Wilk test, which shows that the expression, p=<.001, and the suppression condition, p=<.001, deviated significantly from normality.

Table 2

	Suppression (n=27)	Expression (n=34)
Mean	10.519	8.147
Standard deviation	10.935	10.658
Median	6	5
IQR	7.5	4
Shapiro-Wilk	.757	.578
p-value of Shapiro Wilk	<.001	<.001

Descriptive statistics for thought report 2 split over the instruction condition

Mood Manipulation check

Table 3 summarized the descriptives of the negative and positive mood in phase 1 for the mood induction condition.

Two Welch *t*-test have been conducted to test whether the mood scores for the negative induced condition is significantly different than the neutral induced group in phase 1 for both positive, t(54.098)= -1.409, *p* .164, Cohen's *d* = -.362, and negative, t(58.896)= .320, *p* .750, Cohen's *d* =-.082, mood scores, but no significant effect was found. The results show that there is no significant evidence that participants in the negative mood induced condition will exhibit significantly lower mood ratings compared to those in the neutral mood induced condition as stated in hypothesis 3.

Levene's test for assumption of homogeneity of variances was met for the Welch *t*-test for negative mood score in phase 1, F(1,59)=.728, p=.397, however it was not reached for the Welch *t*-test for positive mood score in phase 1, F(1,59)=4.968, p=.030. The normality of the distribution of the thought report was analyzed by the Shapiro-Wilk test, which shows that the Negative, p=.003, and Neutral condition, p=<.001, deviated significantly from normality in the first phase negative mood scores. However, Negative, p=.762, and Neutral condition, p=.066, does not significantly deviate from normality in the first phase positive mood scores.

Table 3

Descriptive statistics for Negative and Positive mood phase 1 split over the mood induction

condition

	Positive mood score phase 1		Negative mood score phase 1	
	Negative (n=31)	Neutral (<i>n</i> =30)	Negative $(n=31)$	Neutral (<i>n</i> =30)
Mean	25.677	28.433	16.452	15.967
Standard deviation	6.514	8.581	6.136	5.69
Median	25	29	15	14
IQR	7.5	16.5	9	6.75
Shapiro-Wilk	.978	.935	.887	.820
p-value of Shapiro Wilk	.762	.066	.003	<.001

Mood Reinstatement

Table 4 and 5 summarized the descriptives of the second mood phase for both negative and positive mood split for the conditions.

Two ANOVA tests have been conducted to test whether the suppression negative condition differs significantly from the other three conditions in the second phase mood scores, for both positive, F(3,57)=2.306, p=.086, $\eta^2=.108$, and negative, F(3,57)=1.245, p=.302, $\eta^2=.062$, mood scores, but no significant effect was found. Although the effect size does show a medium effect in both positive and negative mood scores in phase two. The results show that there is no significant evidence that participants in the negative suppression condition would exhibit lower mood ratings compared to the three other conditions as stated in hypothesis 4.

The normality of the distribution of the mood score in phase two was analyzed by the Shapiro-Wilk test, which shows that all four conditions do not significantly deviate from normality.

Furthermore, six Welch *t*-tests have been conducted to test that suppression negative is less than the three other conditions for positive and negative mood scores in phase 2. The first two Welch *t*-test have been conducted to test whether suppression negative is significantly less than suppression neutral for both positive, t(15.618)=2.083, p=.973, Cohen's d = .855, and negative, t(13.455)=.997, p=.832, Cohen's d=.421, mood scores, with no significant result found. Levene's test for assumption of homogeneity of variances was met for both positive, F(1,25)=.0014, p=.906, and negative, F(1,25)=.620, p=.439, mood scores in phase 2.

Secondly, two Welch *t*-test have been conducted to test whether expression negative is significantly greater than suppression negative for both positive, t(16.414) = -2.671, p = .992, Cohen's d = -1.035, and negative, t(10.914) = -2.001, p = .965, Cohen's d = -.856, mood scores, with no significant result found. Levene's test for assumption of homogeneity of variances was met for both positive, F(1,29) = .218, p = .644, and negative, F(1,29) = 2.953, p = .097, mood scores in phase 2.

Lastly, two Welch *t*-test have been conducted to test whether expression neutral is significantly greater than suppression negative for both positive, t(18.536)=-1.426, p=.915, Cohen's d = -.622, and negative, t(18.719)=-.926, p=.817, Cohen's d = -.396, mood scores, with no significant result found. Levene's test for assumption of homogeneity of variances was met for both positive, F(1,19)=.559, p=465, and negative, F(1,19)=1.092, p=.310, mood scores in phase 2.

Table 4

	Expression Negative (n=22)	Expression Neutral (<i>n</i> =12)	Suppression Negative (<i>n=9</i>)	Suppression Neutral (n=18)
Mean	20.636	23.667	28.556	22.444
Standard deviation	8.021	8.403	7.265	7.023
Median	19	24	27	22.5
IQR	11.5	11.25	10	12.25
Shapiro-Wilk	.941	.975	.977	.954
p-value of Shapiro Wilk	.206	.956	.945	.487

Descriptive statistics for Positive mood phase 2 split over the conditions

Table 5

Descriptive statistics for Negative mood phase 2 split over the conditions

	Expression Negative (n=22)	Expression Neutral (<i>n</i> =12)	Suppression Negative (<i>n=9</i>)	Suppression Neutral (<i>n</i> =18)
Mean	14.591	15.833	18.889	16.611
Standard deviation	3.887	9.144	5.947	4.828
Median	14	11.5	17	16.5
IQR	5.25	3.75	9	5
Shapiro-Wilk	.889	.609	.927	.941
p-value of Shapiro Wilk	.018	<.001	.456	.303

Individual Research Results

Thought Suppression, QIDS-SR and Positive Mood Change

Table 6 summarized the descriptive of the QIDS-SR scores and the mood change scores.

Pearson's correlation has been conducted to examine the relation between QIDS total score against the positive mood change scores divided between the instruction conditions. The results show that the participants in the suppression condition have a significant strong negative correlation with positive mood change, r = -.524, p = .005. This is visualized in figure 1. The participants in the expression condition show a non-significant weak positive correlation with positive mood change, r = .126, p = .478. This is visualized in figure 2. In the comparison of the correlations a significant difference between suppression and expression on the correlation between QIDS and positive mood change, z = -2.006, p = 0.005 was shown. The results show that there is significant evidence that participants in the suppression condition show a stronger negative correlation than participants in the expression condition with positive mood change.

Figure 1



Linear model of QIDS and Positive Mood Change in Suppression

Note. This figure displays the correlation between the total QIDS score of the participants in the suppression condition and the positive mood change.

Figure 2

Linear model of QIDS and Positive Mood Change in Expression



Note. This figure displays the correlation between the total QIDS score of the participants in the expression condition and the positive mood change.

Thought Suppression, QIDS-SR and Negative Mood Change

Table 6 summarized the descriptive of the QIDS-SR scores and the mood change scores.

Pearson's correlation has been conducted to examine the relation between QIDS total score against the negative mood change scores divided between the instruction conditions. The results show that the participants in the suppression condition have a non-significant strong positive correlation with negative mood change, r=.380, p=.051. This is visualized in figure 3. While the participants in the expression condition show a non-significant weak negative correlation with negative mood change, r=.068, p=.703. This is visualized in figure 4. In the comparison of the correlations a significant difference between suppression and expression on the correlation between QIDS and negative mood change, z=1.722, p=0.043. was shown. The results show that there is significant evidence that participants in the suppression condition show a stronger positive correlation than participants in the expression condition with negative mood change.

Figure 3

Linear model of QIDS and Negative Mood Change in Suppression



Note. This figure displays the correlation between the total QIDS score of the participants in the suppression condition and the negative mood change.

Figure 4

Linear model of QIDS and Negative Mood Change in Expression



Note. This figure displays the correlation between the total QIDS score of the participants in the expression condition and the negative mood change.

Table 6

instruction condition. Condition p-value of Standard Median IQR Shapiro Mean - Wilk Shapiro deviation Wilk QIDS Total 7.618 4.452 6.5 7 Expression .924 .021 (*n*=*3*4) 038 C, 7 667 1 120 6 15 010 Mo 1

Descriptive statistics for QIDS-Total, Mood Change Positive and Negative split over the

I	aD.	le	0	

26

	Suppression $(n=27)$	/.66/	4.438	6	4.5	.919	.038
Mood Change	Expression $(n-34)$	4.765	7.628	6	7.5	.967	.381
Positive	(<i>n</i> -34)						
	Suppression (<i>n</i> =27)	3.259	9.126	2	9.5	.969	.586
Mood Change	Expression $(n=34)$.382	4.314	.5	4.75	.966	.358
Negative	(<i>n</i> -54)						
	Suppression (<i>n</i> =27)	-1.48	6.286	0	5.5	.944	.155

Discussion

"White Bear" Effect and Mood Reinstatement

Contrary to the expectations of hypothesis 1, the study found no statistically significant difference in the first phase between participants in the thought instruction condition regarding the number of target thoughts. This result suggests that thought suppression does not lead to fewer target thoughts compared to expression, indicating a failure to find the thought instruction manipulation. Furthermore, the study found no statistically significant evidence to suggest that participants in the suppression condition reported more target thoughts in the second phase compared to those in the expression condition as stated in hypothesis 2.

Contrary to the expectations of hypothesis 3, the study found no statistically significant difference in the first phase between participants in the mood induction condition regarding the positive and negative mood ratings. This result suggests that negative mood induction does not lead to lower mood ratings, indicating a failure to find the mood manipulation. Secondly, the study found no statistically significant evidence to suggest that participants in the negative suppression condition had lower mood ratings in the second phase compared to the other conditions. This result suggests that thought suppression in combination with negative mood induction does not result in a significant decrease in mood. This result shows no evidence for hypothesis 4 and the mood reinstatement, as negative induced thought suppression does not lead to lower mood ratings.

The original studies (Wegner et al., 1987; Wenzlaff et al., 1991) provided evidence for the paradoxical nature of thought suppression, however subsequent research has produced mixed findings. This study did not replicate the "White Bear" effect and the mood reinstatement (Wenzlaff et al., 1987,1991), contradicting the original results. Instead, this study, due to finding no significant effects, provides support for studies that similarly found no evidence for the "White Bear" effect and mood reinstatement following thought suppression (Merckelbach et al., 1991; Roemer & Borkovec, 1994).

The lack of significant findings in this study suggests the "White Bear" effect and mood reinstatement might not be as robust as proposed in the original studies. However, it is unclear whether the current findings represent a genuine absence of the effect or whether they are the result of insufficient methods to detect a real effect.

A possible explanation for the lack of effect, besides methodological limitations, is that individual differences moderate the outcome (Smari et al., 1994; Rutledge et al., 1993, 1996). This was also highlighted by the original study (Wenzlaff et al., 1991), where they argued that context and psychological factors influence the effectiveness of the "White Bear" effect occurring following thought suppression. This result suggests that the paradoxical nature of thought suppression may not be as universally applicable or as robust as originally stated.

Individual Difference: Depression

In line with the expectations of hypothesis 5, a significant relation was found between depressive severity and positive mood change in those in the suppression condition. Whereby, a higher score on the depressive severity was associated with a negative effect on positive mood change. Suggesting that individuals who engage in thought suppression and have a higher depressive severity will experience a more negative change in positive mood. No significant effect was found in the expression condition. Furthermore, in comparing the correlation between the thought instruction it shows a significant difference between the suppression and expression.

In the examination of hypothesis 6, no significant relations were found between depressive severity, and negative mood change for both thought instruction conditions. However, the suppression condition did show a stronger positive correlation than the expression. Furthermore, in comparing the correlations between the thought instruction it shows a significant difference between the suppression and expression. Suggesting that individuals who engage in thought suppression and have a higher depressive severity will experience a more positive change in negative mood than those in expression.

These findings suggest that the individual difference of depressive severity might moderate the relation between mood change and thought suppression. Furthermore, do these findings align with research in the clinical field, which shows that individuals with higher depressive severity exhibit stronger mood changes when suppressing worrisome thoughts (Silva, 2018; Rosebrock et al., 2019). This suggests that thought suppression may act as a maladaptive strategy, exacerbating mood instability in individuals with depressive symptoms. The work of Wenzlaff et al., (1998) further supports this notion, demonstrating that individuals with a recent depressive episode tend to suppress depressive thoughts as a means of preserving emotional well-being. However, paradoxically, this effort to suppress negative cognition leads to an increased accessibility of such thoughts, ultimately reinforcing the cycle of negative mood and ruminative thinking.

Furthermore, these results show that the interaction between thought suppression and depressive severity also influences a non-clinical population. This aligns with earlier findings by Wenzlaff & Bates (1998), which indicate that individuals with depressive symptoms also struggle to suppress negative thoughts effectively, often experiencing a rebound effect where suppressed content returns with greater intensity. These findings suggest that the processes underlying thought suppression may not be exclusive to a clinical population, but also impact sub-clinical individuals.

Methodological limitations

As stated in the method section, 119 participants are required to reach sufficient statistical power. However, this has not been reached as the study recruited 61 participants.

The lack of statistical power greatly increases the type 2 error inflation (Lakens, 2022). This might be a possible explanation for the cause of non-significant results in the main analysis. Additionally, the smaller overall sample size leads to a smaller sample size in each condition, further reducing the reliability of between-group comparisons and limiting the generalizability of the findings. Secondly, the multiple violated assumptions for the main analysis greatly influences the reliability of the interpretation, as this violation may lead to inaccurate conclusions. In summary, the small sample size and violated assumptions make the result of this study less reliable, robust and accurate.

Furthermore, the mood induction procedure may have also been significantly limited. The music for the mood induction procedure has been adopted from the original study (Wenzlaff et al., 1991). However, several participants have indicated in the feedback that they enjoyed the music, which was intended to induce a negative or neutral mood. In summary, the effectiveness of the mood induction has been called into question.

Lastly, the literature proposes that self-reporting may be a limitation of the study (Wenzlaff et al., 2000), as it may increase the self awareness of the participant and alter their perception, which potentially leads to a heightened awareness (Whetstone & Cross, 1998). This heightened awareness may reduce the likelihood of detecting suppression-related effects, particularly when combined with the demand of the experiment to suppress the target item thought.

Implication further research

Further research can improve the reliability and accuracy of the results by addressing the methodological limitations of this study. Secondly, further research is needed to clarify the conditions under which mood reinstatement may occur, and to determine whether the original findings by Wenzlaff et al., (1991) are generalizable across different populations and contexts.

Practical implications

Regarding practical implications, our results suggest that depressive severity and thought suppression influence mood change scores. This is in line with the idea that thought suppression might be a maladaptive coping strategy that worsens depressive symptoms (Silva, 2018; Rosebrock et al., 2019). Although it is difficult due to our methodological limitations to fully support this claim. The study does support the prominent perspective that treatment options focussing on modifying negative thought suppression might improve mood regulation. According to this view, early intervention on thought suppression might prevent the exacerbation of depressive symptoms, even in a non-clinical population.

Conclusion

In conclusion, the replication study findings demonstrate no support for the "White Bear" effect or the mood reinstatement effect. Whereby, the results support other studies that have not found significant results, indicating the paradoxical nature of thought suppression may not be as robust as originally stated. Whereby the influence of individual differences, might have an impact on the "White Bear" effect. However, the lack of findings may be attributed to methodological limitations, such as sample size and statistical assumptions. Furthermore, it is unclear whether the current findings represent a genuine absence of the effects or whether they are the result of insufficient methods to detect a real effect.

Alternatively, the results did show support for the idea that thought suppression and depressive severity have an influence on positive and negative mood fluctuations. Supporting the idea that thought suppression can be a maladaptive coping strategy, exacerbating depressive symptoms in a non-clinical population.

References

- Clark, D. M., Ball, S., & Pape, D. (1991). An experimental investigation of thought suppression. *Behavior Research and Therapy*, 29, 253-257.
- Clark DM, Winton E, Thynn L. (1993). A further experimental investigation of thought suppression. *Behavior Research and Therapy*, 31:207–10
- Dejonckheere, E., & Bastian, B. (2021). Perceiving social pressure not to feel negative is linked to a more negative self-concept. Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being, 22(2), 667–679 https://doi-org.proxy-ub.rug.nl/10.1007/s10902-020-00246-4
- Eid, M., Gollwitzer, M., & Schmitt, M. (2011). <u>Statistik und Forschungsmethoden Lehrbuch</u>. Weinheim: Beltz.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior Research Methods, Instruments & Computers, 28(1), 1–11. https://doi.org/10.3758/BF03203630
- JASP Team. (2025). JASP (Version 0.19.3) [Computer software]. https://jasp-stats.org/
- Jung, C. G. (2014). Collected Works of C. G. Jung, Volume 9 (Part 1): Archetypes and the Collective Unconscious. *Princeton University Press*.
- Kelly, A. E., & Kahn, J. H. (1994). Effects of suppression of personal intrusive thoughts. Journal of Personality and Social Psychology, 66(6), 998–1006. https://doi.org/10.1037/0022-3514.66.6.998
- Lakens, D. (2022). Improving Your Statistical Inferences (v1.0.0) [Software]. Zenodo. https://doi.org/10.5281/ZENODO.6409077

Lenhard, W. & Lenhard, A. (2014). *Hypothesis Tests for Comparing Correlations*. https://www.psychometrica.de/correlation.html *Psychometrica*. *DOI:10.13140/RG.2.1.2954.1367*

- Merckelbach, H., Muris, P., Van Den Hout, M., & De Jong, P. (1991). Rebound effects of thought suppression: Instruction-dependent? *Behavioural Psychotherapy*, 19(3), 225–238. https://doi.org/10.1017/S0141347300016843
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). https://doi.org/10.1126/science.aac4716
- Purdon, C. (1999). Thought suppression and psychopathology. *Behaviour Research and Therapy*, 37(11), 1029–1054 https://doi-org.proxy-ub.rug.nl/10.1016/S0005-7967(98)00200-9

Qualtrics (2020). Qualtrics (Version November 2024) [Computer software]. Qualtrics.

https://www.qualtrics.com

- Roemer, L., & Borkovec, T. D. (1994). Effects of suppressing thoughts about emotional material. *Journal of Abnormal Psychology*, 103(3), 467–474. https://doi.org/10.1037/0021-843X.103.3.467
- Rosebrock, L. E., Arditte Hall, K. A., Rando, A., Pineles, S. L., & Liverant, G. I. (2019). *Rumination and its relationship with thought suppression in unipolar depression and comorbid PTSD. Cognitive Therapy and Research*, 43(1), 226–235 *https://doi-org.proxy-ub.rug.nl/10.1007/s10608-018-9935-4*
- Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N.,
 Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). "The 16-item Quick Inventory of Depressive

Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): A psychometric evaluation in patients with chronic major depression": Erratum. *Biological Psychiatry*, *54*(5).

- Rutledge, P. C., Hancock, R. A., & Rutledge, J. H. III. (1996). Predictors of thought rebound. Behaviour Research and Therapy, 34(7), 555–562. https://doi.org/10.1016/0005-7967(95)00093-3
- Rutledge, P. C., Hollenberg, D., & Hancock, R. A. (1993). Individual differences in the Wegner rebound effect: Evidence for a moderator variable in thought rebound following thought suppression. *Psychological Reports*, *72*(3), 867–873. https://doi.org/10.2466/pr0.1993.72.3.867
- Shrout, P. E., & Rodgers, J. L. (2018). *Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis*. Annual Review of Psychology, 69, 487-510.
- Silva, S., Janeiro, L., Brás, M., Carmo, C., Martins, A. T., & Jiménez-Ros, A. (2018).
 Paradoxical effects of worrisome thoughts suppression: The influence of depressive mood. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 37(1), 98–106
 https://doi-org.proxy-ub.rug.nl/10.1007/s12144-016-9493-4
- Smari, J., Sigurjonsdottir, H., & Saemundsdottir, I. (1994). Thought suppression and obsession compulsion. *Psychological Reports*, 75(1), 227–235. https://doi.org/10.2466/pr0.1994.75.1.227
- Tucha, L., Fuermaier, A. B. M., Aschenbrenner, S., & Tucha, O. (2013). Vienna test system(VTS): Neuropsychological test battery for the assessment of cognitive functions in

adult ADHD (CFADHD). Schuhfried.

http://www.schuhfried.com/viennatestsystem10/tests-test-sets/all-test-sets-from-a-z/t est/cfadhd-cognitive-functions-adhd-adults/

- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5
- Watson D, Clark LA, Tellegen A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 47 (1063–1070). https://doi.org/10.1037/0022-3514.54.6.1063
- Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality*, 62(4), 615–640. https://doi-org.proxy-ub.rug.nl/10.1111/j.1467-6494.1994.tb00311.x
- Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, 53(1), 5–13. https://doi-org.proxy-ub.rug.nl/10.1037/0022-3514.53.1.5
- Wenzlaff, R. M., & Bates, D. E. (1998). Unmasking a cognitive vulnerability to depression: How lapses in mental control reveal depressive thinking. *Journal of Personality and Social Psychology*, 75(6), 1559–1571. https://doi-org.proxy-ub.rug.nl/10.1037/0022-3514.75.6.1559
- Wenzlaff, R. M., & Wegner, D. M. (1998). The role of mental processes in the failure of inhibition. *Psychological Inquiry*, 9(3), 231–233. https://doi.org/10.1207/s15327965pli0903_17

- Wenzlaff, R. M., Wegner, D. M., & Klein, S. B. (1991). The role of thought suppression in the bonding of thought and mood. *Journal of Personality and Social Psychology*, 60(4), 500–508. https://doi-org.proxy-ub.rug.nl/10.1037/0022-3514.60.4.500
- Wenzlaff, R. M. (1993). The mental control of depression: Psychological obstacles to emotional well-being. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control.* (pp. 239–257). Prentice-Hall, Inc.

World Health Organization. (2022). Mental disorders.

https://www.who.int/news-room/fact-sheets/detail/mental-disorders