Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Suppression-induced forgetting: a pre-registered replication of the think/no-think paradigm

Sera Wiechert, Leonie Loewy, Ineke Wessel, Jonathan M. Fawcett, Gershon Ben-Shakhar, Yoni Pertzov & Bruno Verschuere

To cite this article: Sera Wiechert, Leonie Loewy, Ineke Wessel, Jonathan M. Fawcett, Gershon Ben-Shakhar, Yoni Pertzov & Bruno Verschuere (2023) Suppression-induced forgetting: a pre-registered replication of the think/no-think paradigm, Memory, 31:7, 989-1002, DOI: 10.1080/09658211.2023.2208791

To link to this article: https://doi.org/10.1080/09658211.2023.2208791

Published online: 11 May 2023.

MEMORY
2023, VOL. 31, NO. 7, 989–1002
https://doi.org/10.1080/09658211.2023.2208791

Routledge
Taylor & Francis Group

REGISTERED REPORT

OPEN ACCESS Check for updates

Suppression-induced forgetting: a pre-registered replication of the think/no-think paradigm

Sera Wiechert [a,b], Leonie Loewy [a], Ineke Wessel [c], Jonathan M. Fawcett[d], Gershon Ben-Shakhar[b], Yoni Pertzov[b] and Bruno Verschuere[a]

[a]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands; [b]Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel; [c]Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, The Netherlands; [d]Department of Psychology, Memorial University of Newfoundland, St. John's, Canada

ABSTRACT

Post-traumatic stress disorder is characterised by recurring memories of a traumatic experience despite deliberate attempts to forget (i.e., suppression). The Think/No-Think (TNT) task has been used widely in the laboratory to study suppression-induced forgetting. During the task, participants learn a series of cue-target word pairs. Subsequently, they are presented with a subset of the cue words and are instructed to think (respond items) or not think about the corresponding target (suppression items). Baseline items are not shown during this phase. Successful suppression-induced forgetting is indicated by the reduced recall of suppression compared to baseline items in recall tests using either the same or different cues than originally studied (i.e., same- and independent-probe tests, respectively). The current replication was a pre-registered collaborative effort to evaluate an online experimenter-present version of the paradigm in 150 English-speaking healthy individuals (89 females; $M_{Age} = 31.14$, $SD_{Age} = 7.73$). Overall, we did not replicate the suppression-induced forgetting effect (same-probe: $BF_{01} = 7.84$; $d = 0.03$ [95% CI: −0.13; 0.20]; independent-probe: $BF_{01} = 5.71$; $d = 0.06$ [95% CI: −0.12; 0.24]). These null results should be considered in light of our online implementation of the paradigm. Nevertheless, our findings call into question the robustness of suppression-induced forgetting.

ARTICLE HISTORY
Received 28 June 2021
Accepted 25 April 2023

KEYWORDS
Think/no-think paradigm; suppression-induced forgetting; replication; direct suppression

Changes to the Stage 1 report

- The introduction and method sections are presented in the past instead of future tense, and minor adjustments have been made in line with reviewer's comments.
- The abstract has been updated to reflect our findings and implications.
- The Bayesian stopping rule section (*Sample* in *Methods*) is now rewritten to include sample information of the main study. Therefore, the exclusion criteria were also removed from the *Statistical Analysis* section.
- Pilot data and results have been moved to the OSF repository.
- We added non-pre-registered exploratory analyses (and we specified that the analyses were exploratory) at the end of the *Results* section.
- Minor spelling and grammar errors in the introduction and method sections were corrected. The original Stage 1 report can be found on OSF at https://osf.io/jnk78/.

Throughout our lives, we accumulate a variety of unpleasant memories. In the case of particularly negative and uncontrollable events such as assault, physical or psychological violence, these memories are not only aversive but can contribute to the development of post-traumatic stress disorder (PTSD). PTSD is characterised by recurring, intrusive thoughts that risk detracting from everyday function (American Psychiatric Association, 2013). Patients are often motivated to avoid such intrusions, with the eventual goal of preventing the traumatic event from coming to mind as often or at all (e.g., Horowitz, 1976; Williams et al., 1999). However, this commonly observed behaviour in individuals with PTSD raises the question of whether a deliberate cognitive act, such as the wilful suppression of a memory, can have lasting memory-impairing effects.

Suppression-induced forgetting refers to the phenomenon that preventing memories from entering awareness by actively suppressing them impairs later recall (Hertel & McDaniel, 2010; Stramaccia et al., 2021). To study this phenomenon, Anderson and Green (2001) introduced the Think/No-think (TNT) paradigm, which typically

CONTACT Sera Wiechert s.wiechert@uva.nl Department of Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 WT Amsterdam, The Netherlands

consists of three main phases (e.g., Anderson et al., 2004; Anderson & Green, 2001). First, participants study stimulus pairs (i.e., cue-target pairs; e.g., WAFFLE-MAPLE) until a minimum number of pairs are retained (e.g., 50% of all word pairs in the original study by Anderson & Green, 2001). Second, participants are presented with a series of individual cues (e.g., WAFFLE) and are instructed to either think (i.e., respond condition) or not think about the target (i.e., MAPLE; suppression condition) of the cue-target pair. Other items that are assigned to the baseline condition are not presented at all during this phase. Third, participants are then tested for their memory of the target word in either or both of two separate tests (same-probe, independent-probe). In the same-probe test, the cue word is presented (e.g., WAFFLE), and participants must try to recall the target item of the pair (i.e., MAPLE). In the independent-probe test, they are presented with an item that is semantically associated with the target along with the first letter of the target (e.g., TREE-M) and are instructed to recall the target item of the pair (i.e., MAPLE; see Figure 1). Reduced recall of no-think items compared to baseline items provides evidence of suppression-induced forgetting.

The independent-probe test has been viewed as particularly important when it comes to inferring the origins of the underlying mechanism of suppression-induced forgetting (Anderson & Levy, 2007). While the same-probe test effect may also be explained by reduced associative strength between the original cue and its target (or related interference effects), the independent-probe test effect is more specifically related to the accessibility of the target word itself (Anderson & Green, 2001; Anderson & Huddleston, 2012; Bulevich et al., 2006; however, see also Racsmány et al., 2012 for alternative explanations). If stopping retrieval would result in inhibition in the sense of a decreased activation of the target's memory trace, then the suppression-induced forgetting effect should also generalise to independent probes (Anderson & Levy, 2007). In contrast, if suppression-induced forgetting is due to associative interference (e.g., interference with competing thoughts), it would not generalise to independent probes. As we are interested in distinguishing between interference and inhibition as potential mechanisms to explain suppression-induced forgetting (Anderson & Levy, 2007), both memory tests were included in the current effort.

The TNT task has received much attention within the scientific literature. According to Google Scholar, the first study introducing the paradigm (i.e., Anderson & Green, 2001) has been cited over 1400 times (Web of Science: 675 citations, April 2023). Additionally, examining suppression-induced forgetting is not only relevant from a theoretical standpoint but has clinical and forensic relevance (e.g., Hertel & Mahan, 2008; Hertel & McDaniel, 2010; Joormann et al., 2009; Waldhauser et al., 2018). The original study on the TNT paradigm (Anderson & Green, 2001) suggested that the mechanisms behind the TNT effect provide "a viable model for repression" (p. 366; also see Conway, 2001). The comparison with repression suggests that suppression-induced forgetting bears relevance to the repressed memory debate that arose in the 1990s (i.e., "memory wars"; Patihis et al., 2014). In short, the centre of this debate is the notion that traumatic memories exist in an unconscious form, causing psychopathological symptoms until they are made accessible through memory-recovery techniques. Despite this debate, the belief in repressed memory (Otgaar et al., 2021) and in the utility of memory-recovering techniques persists (Dodier et al., 2021; Otgaar et al., 2019). Therefore, accumulating evidence on suppression-induced forgetting within the TNT paradigm may further characterise the nuances inherent in these ongoing debates.

Multiple studies found statistically significant suppression-induced forgetting (e.g., Anderson et al., 2004; Anderson & Green, 2001; del Prete et al., 2015; Joormann et al., 2005; Murray et al., 2011; Taubenfeld et al., 2019). Yet, others report mixed results (e.g., Meier et al., 2011; Noreen & MacLeod, 2014; Racsmány et al., 2012), null findings (e.g., Bulevich et al., 2006), or that the statistical significance depends on data-analytic decisions (Wessel et al., 2020). To our knowledge, there have been two published attempts to summarise the literature on the TNT evidence until the submission of our Stage 1 report. However, the one published systematic review on the TNT effect has focused on specific group comparisons (i.e., clinical vs healthy groups in Stramaccia et al., 2021), and the other review has not been systematic or exhaustive (i.e., Anderson & Huddleston, 2012).

In the meta-analysis focusing on clinical samples (i.e., individuals affected by a range of internalising and externalising psychological disorders or exhibiting high scores on related traits), Stramaccia et al. (2021) found an average same-probe effect of 0.31 (Hedges' g; 95% CI [0.16, 0.45]). However, only very few studies relevant to Stramaccia et al.'s (2021) clinical focus reported independent-probe results, which was insufficient for conducting a meta-analysis and retrieving reliable average effect size estimates (Stramaccia et al., 2021). There may be several explanations for the relative paucity of independent-probe results. First, the independent-probe effect may have simply been studied less often. In some cases, independent-probe testing is less relevant for the type of questions being asked (i.e., tests of interference rather than inhibition; e.g., the effect of cue relatedness in dysphoric and non-dysphoric participants, Hertel & Mahan, 2008). In other cases, it may be difficult to devise an independent cue for the stimulus material that is employed (e.g., aversive pictures, Depue et al., 2006; autobiographical memories, Noreen & Macleod, 2013). Alternatively, as one of us speculated previously (Wessel et al., 2020), there may be a true independent-probe effect, but its size may be smaller than that of the same-probe effect. If this is the case, independent-probe studies basing their sample size calculations on same-probe effect sizes may lack

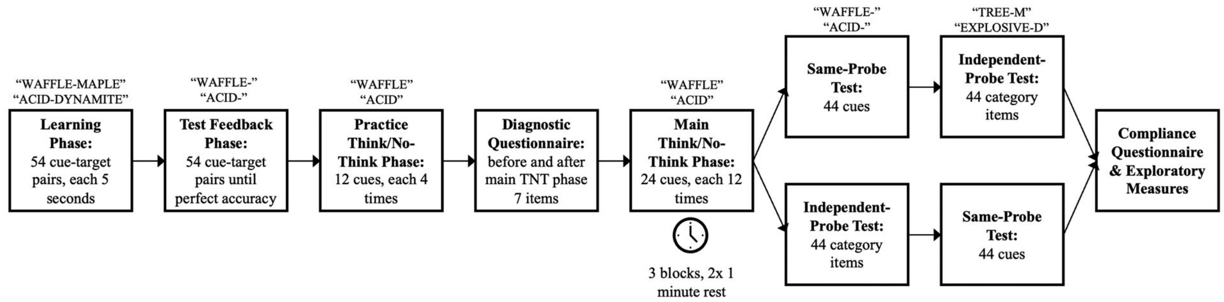

Figure 1. Overview of the procedure flow.

sufficient statistical power to show a statistically significant finding (i.e., Type 2 error). As null findings are less likely to be published (i.e., publication bias, Rosenthal, 1979), fewer independent-probe results may be available in the literature. We note that we do not know to what extent the TNT literature suffers from publication bias. Yet, Stramaccia et al.'s (2021) meta-analysis indicated that, for the same-probe effect in healthy control groups, some studies with findings within the statistically non-significant range might be missing from the literature. We are presently unaware of any other, more exhaustive meta-analyses published in the literature to shed more light on this issue, although at least one is pre-registered (i.e., https://osf.io/gdm79) and available online in the form of a master thesis project (Clark, 2021).

Even though the above deliberation on the TNT evidence (specifically, the independent-probe results) may appear to only indicate a predominant need for meta-analytic work, we argue that larger pre-registered replications are also necessary. Meta-analyses can be a valuable step in evaluating the literature, but they are not a viable substitute for pre-registered replications (Van Elk et al., 2015). Pre-registered replications can counteract potential problems of experimenter or publication bias (Van Elk et al., 2015). As we do not know the extent to which the TNT literature suffers from potential biases, and meta-analytic work is already underway, we argue that a well-powered pre-registered replication is vital at this stage.

Table 1. Overview of differential outcome possibilities for both replication outcomes.

	Same-probe successful	Same-probe unsuccessful
Independent-Probe successful	Evidence for suppression-induced forgetting. TNT successfully decreased the accessibility of the target word as well as the associative strength of the cue-target association	Partial evidence for suppression-induced forgetting. TNT successfully decreased the independent accessibility of the target word, but did not alter the associative strength of the cue-target association
Independent-Probe unsuccessful	Partial evidence for suppression-induced forgetting. TNT successfully weakened the cue-target association, but did not alter the independent accessibility of the target word	No evidence for suppression-induced forgetting

Therefore, to contribute to the accumulating literature and facilitate more reliable suppression-induced forgetting effect size estimates for both the same- and independent-probe tests, we report here a pre-registered replication of the TNT paradigm. For an overview of the key elements of our replication, please see Table A1. Here, rather than replicating one published protocol, we applied a paradigmatic replication (inspired by, but not identical to the replication format used by Vohs et al., 2021). A direct replication aims to repeat a specific published protocol to produce the effect of interest (Simons, 2014). However, a downside of direct replications is that the original protocol may not incorporate subsequent knowledge on the effect of interest. To mitigate this, we chose to conduct a paradigmatic replication. Based on a review of the literature, we (SW, BV, YP, GBS) developed a protocol through consultation with researchers who have published relevant studies on suppression-induced forgetting (IW, JF), see the Appendix.

Two design choices in our paradigmatic replication deserve special attention. First, the current study was implemented online. To this end, we created a virtual experimenter-present version to ensure instruction comprehension and adherence, motivation to actively suppress and an adequate environment to facilitate concentration. Second, we implemented the direct suppression instruction (i.e., suppress the target item while not generating distractive thoughts; e.g., Benoit & Anderson, 2012) rather than thought substitution (i.e., avoid the target stimuli by thinking of alternative thoughts; e.g., Hertel & Calcaterra, 2005). In line with the reasoning of Anderson and Levy (2007), only direct suppression may trigger an effect in both same- and independent-probe measures, as substituting a target with an alternative item may weaken the cue-target association (same-probe test) by creating associative interference, but not necessarily inducing active target word inhibition (independent-probe test). Thus, to align our instructions with both memory tests and be able to examine a reliable suppression-induced forgetting effect, direct suppression instructions were used.

Taken together, the present study aimed to replicate the TNT task in a highly powered online experimenter-present version using a direct suppression instruction. We assessed the suppression-induced forgetting effect in both the same- and the independent-probe tests, see Table 1. Hence, we examined whether individuals show reduced correct recall for target words in the no-think condition compared to the baseline condition for both tests.

Method

Ethics approval was obtained by the Ethics Review Board of the University of Amsterdam (2021-CP-13319). Data collection and the Stage 2 submission both occurred in the pre-registered time frame.

Data availability statement

All materials, scripts and data that support the findings of this study are openly available on the Open Science Repository at https://osf.io/e75a6/.

Sample

According to the pre-registered stopping rule, we stopped data collection when we obtained substantial evidence for either hypothesis in both analyses (same-probe and independent-probe tests show BFs > 5) at 150 inclusions (i.e., after the third batch in a sequential inclusion procedure). In total, 213 participants started the session after verifying the experimental control questionnaire. Seven out of the 213 participants did not complete the session (3.29%) because they either wanted to stop, left the session meeting unexpectedly or experienced unforeseen technical problems (e.g., loss of internet connection). Further, 39 participants (18.31% of $n = 213$) timed out in the test feedback phase (i.e., failed to learn all word pairs within the 25-minute time limit; please see the *Test Feedback* in the *Procedure* section for more detailed information on the drop-off procedure during learning). Thus, 167 participants completed the session. Of those 167 completers, two participants (1.20%) had to be excluded because they experienced long internet glitches at critical phases (one in the TNT phase and one in the same-probe and independent-probe tests). Moreover, 15 participants (8.98% of $n = 167$) were excluded because they scored above 4 in the compliance screening. We did not exclude any participant for the pre-registered criteria of accurately reacting to at least 75% of the think and no-think trials during the main phase (i.e., no target recall for red cues, correct target recall for green cues). Additionally, no participant indicated "very much" on either distraction questions (Questions 2 and 10) of the session evaluation questionnaire (Zwaan et al., 2018). Thus, 150 participants (70.42% of 213 session beginners; 89.82% of 167 session completers) were included in the analyses. The final sample included 89 females (61 males, $n = 150$), and participants had a mean age of 31.14 ($SD = 7.73$, $n = 150$). Further, the sample comprised of 75 Western (50% of $n = 150$; 46 females, 29 males; M age = 33.80, SD age = 8.03) and 75 non-Western participants (50% of $n = 150$; 43 females, 32 males; M age = 28.48, SD age = 6.44).

Data were collected through the online portal *Prolific.co*, with experimenters from collaborating laboratories (i.e., The Netherlands, Canada and Israel). Participants completed one session (approximately 75 min) and were compensated monetarily (approximately 12 euros; partial payments were possible when the experimental session stopped early, e.g., due to a learning phase time-out). To make use of the online format of the study and increase generalisability, the current study included half of the participants from Western and half from non-Western countries (based on nationality) in which (one of) the main languages is English (i.e., primary, de facto or de jure language; e.g., US and South-Africa, respectively). The inclusion criteria for study participation were: language fluency in English, current physical and psychological health (i.e., as this can confound the TNT effect; Stramaccia et al., 2021), normal or corrected-to-normal vision, age between 18 and 45 (i.e., to minimise developmental differences; Paz-Alonso et al., 2009), and willingness to participate in a video call session. These inclusion criteria were built-in pre-screening measures of Prolific and we filtered participants according to our criteria (i.e., no data collection on inclusion measures). In addition, the session only commenced if accurate online data collection was possible: participants needed to have access to the video communication software, use a laptop or pc (i.e., with an integrated camera), ensure a stable internet connection, and sit in a calm place of their home with no distractions. Adherence to these criteria was verified using the experimental control questionnaire at the beginning of the session (see *Materials*). The session only commenced if participants' responses to this questionnaire were satisfactory.

Design

The experimental design included three levels of the word pair condition (baseline vs think vs no-think) as the within-subject independent factor. The dependent variables were the percentage of correct recall in the same- and independent-probe tests.

Materials

Stimuli

The TNT study included 54 cue-target pairs (e.g., WAFFLE-MAPLE), divided into three groups of 12-word pairs and one of 18 filler pairs. The three lists of 12-word pairs were counterbalanced across no-think, think and baseline conditions (e.g., our sample of 150 participants included 50 participants in each of the counterbalancing conditions).

The stimuli set by Benoit and Anderson (2012) was used (see in full on https://osf.io/e75a6/).

Experimental control questionnaire

The experimenter went through the experimental control questionnaire at the beginning of the session to assess whether the technical setup was working and the environmental situation was sufficiently adequate to ensure an appropriate testing environment. The questionnaire (in full on https://osf.io/e75a6/) consisted of 11 items, including technical setup questions such as: "Has the participant switched off their phone?" and environmental condition questions such as: "Is it noisy in the background of the participant?". Only if the technical setup was complete and the environmental condition adequate could the session commence.

Diagnostic questionnaire

We used the diagnostic questionnaire (as in Anderson et al., 2004) to ensure that participants understood and accurately followed the instructions. The questionnaire (in full on https://osf.io/e75a6/) consisted of seven items, including questions such as: "For the green hint words, how often did you try to come up with the associated response as fast as possible?". Responses to this questionnaire were not used for confirmatory analyses but rather to clarify the instructions during the session (see *Procedure*).

Compliance questionnaire

The compliance questionnaire (as used in, e.g., Levy & Anderson, 2012; van Schie et al., 2013; van Schie & Anderson, 2017; in full on https://osf.io/e75a6/) was used to assess whether participants complied with the instructions throughout the main TNT phase; it included three statements such as: "I read the red cue word, tried to not think of the associated response, but then after the trial was over, I made sure that I still remembered the target word" [0 – Never; 4 – Very frequently; range total score: 0–12]. Participants with a total score above 4 (out of a total of 12) were excluded and replaced according to their list counterbalancing conditions.

Session evaluation questionnaire

We included a session evaluation questionnaire as used in Zwaan et al. (2018) at the end of the session (in full at https://osf.io/e75a6); this questionnaire included nine items such as "All instructions were clear", measured on a 3-point Likert scale [1 – Not at all; 3 – Very much]. For the exclusion criteria, if participants selected "Very much" on either of the distraction questions in the questionnaire (i.e., "There are a lot of distractions here"; "I was distracted during the experiment"), they were excluded from the analyses.

Demographics

To describe our sample, we asked participants to indicate their gender [Male; Female; Non-binary/third gender] and age [open text entry], each with an additional "Prefer not to say" option.

Lextale

To validate English proficiency in experimenters, we used the English version of the LexTale test by Lemhöfer and Broersma (2012). This is a brief vocabulary test consisting of 60 stimuli, presented one by one on the computer screen, in which the task of the participant is to judge whether the presented stimulus is an existing word or not. The test consisted of 20 pseudo- (e.g., crumper) and 40 existing words (e.g., savoury). The LexTale score is the percentage of correctly judged stimuli, calculated as ((number of words correct/40*100) + (number of non-words correct/20*100)) /2. In the study by Hendriks et al. (2021), native English speakers had an average score of 89% ($SD = 11.64$), while Dutch and other international advanced English learners scored on average 76% $SD = 11.85$). Further, the study by Lemhöfer and Broersma (2012) has shown that LexTale scores correlate substantially with general English proficiency and are superior in predicting proficiency compared to self-ratings.

Exploratory measures

For exploratory purposes, additional brief measures (i.e., sleep quality and quantity, tiredness, hunger and thirst, executive function problems, perceived stress, substitution use, belief in suppression and repression, and thought control ability) were also assessed (full description on https://osf.io/e75a6/).

Procedure

The study was conducted in English, and our recruitment pool focused on fluent English speakers. However, to make use of the online format of the study and increase generalisability, participants from multiple native English-speaking countries were included (50% Western, 50% non-Western). Further, to ensure that all experimenters were competent to test in English, they were required to take the LexTale test (Lemhöfer & Broersma, 2012) before testing and needed to score at least 80% (i.e., C1 & C2 level based on Lemhöfer & Broersma, 2012) to be able to qualify for the experimenter role. Additionally, experimenters were only qualified to test if they knew all word pairs with 100% accuracy and passed the mock trial sessions (i.e., example testing sessions with a mock participant; with pre-specified scripts and scenarios). To pass the mock trial sessions, experimenters were extensively trained by a senior experimenter across multiple sessions (i.e., introductory meetings and mock trial sessions). During the mock trial sessions, experimenters led two trial sessions and were evaluated by senior experimenters. Experimenters were only verified to test for the current study if they sufficiently adhered to the response scripts and procedures. The detailed experimenter verification procedure can be found at https://osf.io/e75a6/.

Additionally, throughout the testing period, senior experimenters monitored and re-evaluated the adequacy of the experimenter's sessions and gave feedback to the experimenter after each 15th to 20th testing session.

For a full script of the instructions, see https://osf.io/e75a6/. Participants booked a video call appointment with the experimenter after consenting to participate in the session. The experimenters ran the programme on their computer and ensured that the participant saw the shared screen (but not the video window). The experimenter coded all responses based on the participants' recalled answers. Participants were told that the study was about attention and that its goal was to measure individuals' capacity to avoid distraction. The TNT task entailed seven phases; at the end of the task, participants received additional end-of-study questionnaires (see Figure 1 for an overview).

Learning phase

In this phase, 54 cue-target pairs (i.e., 18 filler pairs; 12 baseline pairs, 12 think and 12 no-think pairs counterbalanced across participants) were presented once in the middle of the screen for 5 seconds (white font on a black background; in semi-random order: first and last trials consisting of filler items and the other items randomised in between), with the target word displayed to the right of the cue. Participants were instructed to learn the cue-target word pairs in preparation for a later unspecified test.

Test feedback

After the learning phase, participants' memory of the cue-target pairs was tested. This time only the cues were presented in the middle of the screen for 5 seconds, and participants were asked to recall the associated target as fast as possible. We used a drop-off procedure (Levy & Anderson, 2012). When participants correctly recalled the target upon seeing the cue, this pair was dropped off the list of the to-be-learned words (i.e., indicated by the researcher's button press). If participants failed to recall the correct associated target, the correct answer was presented in the middle of the screen in blue for 2.5 seconds and remained in the to-be-learned words; thus, it was tested later on until accurately recalled. This process was repeated until participants reached 100% accuracy for the word pairs or if the time limit of 25 minutes was reached (i.e., time-out exclusion).

Practice think/no-Think phase

Once participants completed the test feedback phase, they were instructed on the TNT phase (both on the screen and verbally). In line with Benoit and Anderson (2012) and Taubenfeld et al. (2019), the experimenter stressed that participants should suppress the retrieval of no-think targets while also not generating alternative thoughts. In the practice phase, cues were presented in the middle of the screen for 3.5 s. Think cues appeared in green and indicated that participants needed to recall the associated target aloud, whereas no-think cues appeared in red, for which participants were instructed not to recall nor think about the associated target. Additionally, we included feedback in the TNT phase. When think cues appeared in green and participants did not recall the associated response within the time frame, the associated target appeared in blue for 2 seconds. In contrast, when no-think cues appeared in red, and participants recalled the associated target, an error message appeared. The practice phase included 48 trials with 12 filler pairs, in which both the six no-think and think cues were presented four times each.

Diagnostic questionnaire

After the practice TNT phase and in the middle of the main TNT phase, the diagnostic questionnaire appeared on the screen. The experimenter and participant went through the questionnaire together and used this phase as an interactive process to repeat and clarify TNT instructions, hereby tailoring parts of the instructions to the responses given by the participant.

Think/no-think phase

This phase entailed the same colour-cueing and feedback procedure as in the practice TNT phase. 24 cues (12 no-think, 12 think) were presented in the middle of the screen in a random intermixed order. All cues were presented 12 times each, resulting in 288 critical trials (144 no-think and 144 think trials). Most studies that used 12 or 16 repetitions as their highest condition have shown statistically significant effects (Anderson & Huddleston, 2012). We decided on 12 repetitions to extract a robust suppression-induced forgetting effect while guarding against fatigue effects that may arise from 16 repetitions. Trials were divided into three blocks of 96 trials with a 1-minute rest period in between.

Same- and independent-probe test

In this stage, participants were tested on their target memory in same- and independent-probe tests while randomising their order across participants. Participants were instructed to respond to all cues irrespective of whether they were associated with green or red words in prior phases. The experimenter coded the recalled responses on the keyboard. In the same-probe test, the cue was presented in the middle of the screen (e.g., WAFFLE), and participants were asked to recall out loud the associated target (e.g., MAPLE). In the independent-probe test, participants were asked to recall all targets upon seeing the word associated with the target and the first letter of the associated target (e.g., TREE-M). Each test started with eight filler items, followed by the 36 critical items. Critical items in both the same-probe and the independent-probe test were presented in random order.

Compliance screening and end-of-Study questionnaires

For both the compliance screening and the end-of-study questionnaires, a Qualtrics link was sent via the video chat function. Once participants completed the questionnaire, they were asked to share their screens. Then, participants were thanked for their participation, debriefed and compensated. Debriefing occurred verbally, and participants were also able to download the debriefing document.

Statistical analysis

Fully anonymised raw data was made publicly available upon publication. All analyses were conducted on the final sample of $n = 150$ after listwise exclusions. For all confirmatory analyses, we used a Bayesian approach. Data preparation and analyses were conducted in RStudio (Version 1.2.5019) and JASP (Version 0.14.1; for the Bayesian analyses). For the Bayesian analysis, a standard Cauchy JZS prior with scaling factor r = .707 was used. We calculated Bayes Factors (BF_{01}), which indicate how much more likely the data are under the null hypothesis compared to the alternative hypothesis (Jarosz & Wiley, 2014). The inverse ratio allows one to speak to the likelihood of the alternative hypothesis compared to the null (Jarosz & Wiley, 2014).

Manipulation check

We tested for a positive control effect in the same-probe test in our final sample; thus, we compared the null hypothesis (i.e., no difference in correct recall of targets between the think and the baseline condition) to a one-sided alternative (i.e., correct recall of targets in the think condition is *higher* than that of the baseline condition). This *facilitation effect* reflects pure practice effects of think targets in the think/no-think phase and has even been replicated in studies that showed null results for suppression-induced forgetting (e.g., Bulevich et al., 2006). To this end, we conducted a Bayesian one-tailed student paired samples t-test comparing the percentage of correct recall of targets in the think and baseline groups. As pre-registered, before testing the facilitation effect, we removed extreme outliers (i.e., difference score that is larger or smaller than the third and first quartiles, respectively, by three times of the interquartile range). Due to using a large sample size in this study, we refrained from testing the normality assumption. Further, we investigated our data for a ceiling (mean baseline accuracy > 95%) or floor effect (mean baseline accuracy < 20%) in both tests.

Confirmatory analysis

The first outcome of interest was the suppression-induced forgetting effect in the same-probe test; thus, we conducted a Bayesian one-tailed student paired samples t-test to compare the null hypothesis (i.e., no difference in correct recall of targets between the no-think and the baseline condition) to a one-sided alternative (i.e., correct recall of targets in the no-think condition is *lower* than that of the baseline condition). For the second outcome, namely suppression-induced forgetting effect in the independent-probe test, we equally conducted a Bayesian one-tailed student paired samples t-test to compare the null hypothesis (i.e., no difference in correct recall of targets between the no-think and the baseline condition) to a one-sided alternative (i.e., correct recall of targets in the no-think condition is *lower* than that of the baseline condition). As mentioned above, before conducting confirmatory analyses on both outcomes, we removed extreme outliers per analysis. We report *Cohen's d* effect sizes with their respective 95% confidence intervals.

As pre-registered, replication success for each outcome was classified as follows: if the Bayesian analysis showed $BF_{10} > 5$, we would conclude that the replication of the respective effect conveyed convincing evidence. If we found $BF_{10} < 0.2$ ($BF_{01} > 5$), we would report the Bayesian analysis to show convincing effects in favour of the null hypothesis over the alternative hypothesis for the respective effect. Finally, if the Bayesian analysis would show $BF_{10} < 5$ and $BF_{10} > 0.2$ ($BF_{01} > 5$), we reported that we did not successfully replicate the respective effect and could not draw definite conclusions on its directionality.

Results[1]

Manipulation checks

There was no ceiling or floor effect for baseline recall in the same-probe ($M = 79.17\%$, $SD = 15.89\%$) or the independent-probe test ($M = 73.89\%$, $SD = 17.32\%$). Further, our positive control was successful as participants correctly recalled more words in the think condition ($M = 97.78\%$, $SD = 4.80\%$) than in the baseline condition ($M = 79.17\%$, $SD = 15.89\%$; $BF_{10} = 3.89e^{+28}$; $d = 1.46$ [95% CI: 1.18; 1.74]).

Confirmatory analyses

To examine the suppression-induced forgetting effect in the same-probe test, we conducted a Bayesian one-tailed student paired samples t-test. One extreme outlier (i.e., a difference score of 75% between the baseline and suppression condition) was excluded from the analysis. The results showed conclusive evidence in favour of the null over the alternative hypothesis ($BF_{01} = 7.84$; $d = 0.03$ [95% CI: −0.13; 0.20]) as such that the data are about eight times more likely under the null compared to the alternative hypothesis. Thus, we did not find conclusive evidence for lower memory performance in the suppression condition ($M = 78.58\%$, $SD = 18.54\%$) compared to the baseline condition ($M = 79.14\%$, $SD = 15.94\%$; think condition: $M = 97.78\%$, $SD = 4.80\%$).

Similar results emerged for the independent-probe test. Again, the Bayesian one-tailed student paired samples t-test showed conclusive evidence in favour of the null over the alternative hypothesis ($BF_{01} = 5.71$; $d = 0.06$ [95% CI: −0.12; 0.24]). The independent-probe test also did not indicate conclusive evidence for lower accurate recall of suppression ($M = 72.83\%$, $SD = 15.67\%$) compared to baseline items ($M = 73.89\%$, $SD = 17.32\%$; think condition: $M = 79.22\%$, $SD = 15.74\%$). For the independent-probe test, the data are about six times more likely under the null than the alternative hypothesis.

Exploratory analyses

We conducted additional analyses to investigate the robustness of the results. First, we re-examined the same-probe and independent-probe results when only including participants with perfect self-reported compliance (score of 0 in the compliance screening, $N = 63$). When re-analysing this, both the same-probe ($BF_{01} = 5.11$; $d = 0.05$ [95% CI: −0.19; 0.29]) as well as the independent-probe test ($BF_{01} = 4.46$; $d = 0.08$ [95% CI: −0.21; 0.36]) still showed evidence in favour of the null over the alternative hypothesis, but inconclusive for the independent-probe outcome.

Second, as the same- and independent-probe tests were presented in randomised order, practice effects may have obscured the suppression-induced forgetting effect. To examine this possibility, we re-analysed our data with only participants who first completed the respective test. When analysing the same-probe effect only in participants who received the same-probe test first ($N = 69$), the results still tended to favour the null hypothesis ($BF_{01} = 2.83$, $d = 0.12$ [95% CI: −0.12; 0.37]), although inconclusive. When re-analysing the independent-probe hypothesis on first completers only ($N = 81$), the Bayes factor showed conclusive evidence in favour of the null over the alternative hypothesis ($BF_{01} = 10.62$, $d = -0.05$ [95% CI: −0.31; 0.22]).

Third, we examined whether the same- and independent-probe effects differed per counterbalancing condition of the word pairs ($N = 50$ in each of the three counterbalancing conditions). For this, we conducted a Bayesian repeated-measures ANOVA, including the repeated-measures factor (suppression vs baseline condition) and the counterbalancing variable (Condition A vs B vs C) and their interaction. For the same-probe test, we found that the null model best explained our data, and none of the effects reached the threshold for inclusion (word pair condition: $BF_{Incl} = 0.25$, $\eta^2 = 8.94e^{-4}$, counterbalancing condition: $BF_{Incl} = 0.47$, $\eta^2 = 0.02$, interaction: $BF_{Incl} = 0.76$, $\eta^2 = 0.02$). In contrast, for the independent-probe test, the best model explaining the data was inclusive of both predictors and their interaction ($BF_{10} = 495.21$, compared to the null model; interaction: $BF_{Incl} = 1295.33$, $\eta^2 = 0.05$). As shown in Figure 2, while counterbalancing conditions A and B showed no evidence of suppression-induced forgetting, condition C displays a pattern that is in line with suppression-induced forgetting (suppression: $M = 0.71$, $SD = 0.17$, baseline: $M = 0.81$, $SD = 0.18$).

Lastly, we investigated whether the same- and independent-probe effects differed per sample employed (75 in the Western and non-Western groups). Similarly, we conducted a Bayesian repeated-measures ANOVA, including the repeated-measures factor (suppression vs baseline condition) and the sample variable (Western vs non-Western) and their interaction. For the same-probe test, we found that the best model explaining the result included the sample variable only ($BF_{10} = 5.65$, compared to the null model), with a BF_{Incl} of 3.86 ($\eta^2 = 0.04$). Hence, there was a main effect of the sample indicating that accuracy did not differ between suppression and baseline items in both groups, but the Western sample displayed higher accuracy overall (see Figure 3). For the independent-probe test, the best model explaining the effects included just the sample category ($BF_{10} = 15188.21$, compared to the null model; BF_{Incl} of 10447.17, $\eta^2 = 0.10$). However, the model including the interaction between word pair condition (suppression vs baseline) and sample category also yielded a conclusive Bayes factor ($BF_{10} = 592.02$, compared to the null model). Plotting the result (see Figure 4) demonstrates that the Western sample showed higher overall accuracy. Additionally, although inconclusive ($BF_{Incl} = 0.14$), the Western sample showed a slight tendency towards suppression-induced forgetting (i.e., decreased accuracy for suppression items), whereas the non-Western sample showed no such pattern.

To explore further associations between the same-probe effect and the independent-probe effect on the one hand, and other relevant variables to the study on the other hand (i.e., TNT-related measures, exploratory measures), we conducted Bayesian Pearson's bivariate correlation analyses (see Table 2). Importantly, the compliance score was not associated with the same-probe ($r(150) < .01$, [95% CI: −0.16; 0.16], $BF_{10} = 0.10$) or the independent-probe difference score ($r(150) = .01$, [95% CI: −0.15; 0.17], $BF_{10} = 0.10$) between the suppression and baseline condition. In addition, the same- and independent-probe effects were not associated with any other exploratory measure, i.e., substitution use in the TNT phase, executive functioning problems, perceived stress in the past month, the session evaluation or thought control ability (see Table 2).

Discussion

The current study sought to replicate the suppression-induced forgetting effect in a highly powered study, using an online experimenter-present version of the TNT task with the direct suppression instruction. Our positive control – i.e., increased memory accuracy for think compared to baseline items – was successful. As there were also no ceiling or floor effects in the same- and independent-probe tests, we could readily interpret the confirmatory analyses. Overall, our participants did not show reduced accurate recall in the suppression compared to the baseline condition. Thus, we did not find

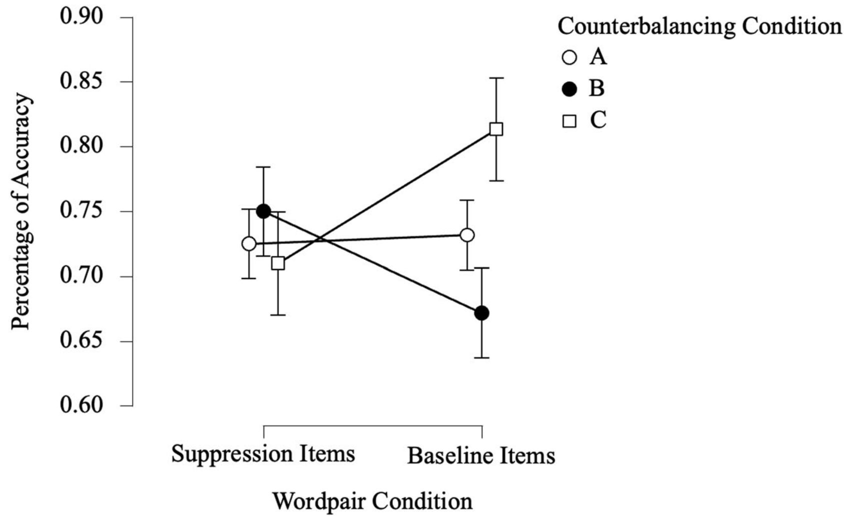

Figure 2. Differences between the independent-probe suppression condition average accuracy and baseline condition average accuracy divided by the counterbalancing condition with error bars representing the 95% confidence interval.

suppression-induced forgetting in the same- and independent-probe tests.

Notably, our sensitivity analyses testing the robustness of our confirmatory results revealed two interesting results. First, we found higher overall memory accuracy for both the baseline and suppression condition in the same- and independent-probe test for Western participants (vs non-Western). However, because we did not measure other memory-related measures (e.g., highest obtained education), we cannot pinpoint potential differences in this respect.

Second, we obtained differential suppression-induced forgetting effects in the independent-probe test depending on the list counterbalancing condition; we observed successful suppression-induced forgetting in the list condition C, but a reversed effect in B. Further inspection showed that the discrepancy across list conditions was mainly attributed to differing average baseline item memory per list condition, while suppression item memory was similar. As such, item-specific effects could be speculated (e.g., specific word groups being generally more difficult or easier to memorise), which might have biased the observed null results. Yet, if this was the case, the same pattern of list condition differences would be expected in the same-probe test; interestingly, no such pattern was observed. Hence, even considering only specific list conditions does not lead to reliable suppression-induced forgetting in both tests. Additionally, as post-hoc sensitivity analyses are exploratory and speculative in nature, and the statistical power of subgroup analyses is reduced, sensitivity analyses should be treated cautiously.

This replication study contributes to the accumulating literature and facilitates more reliable suppression-induced forgetting effect size estimates for both the same- and independent-probe tests. We endeavoured to ensure study quality and, thus, increase confidence in our findings. That is, rather than reproducing the original or most cited study, we chose to conduct a paradigmatic replication style in which we combined design elements that have proven insightful and valuable in the literature. Further, in order to minimise researcher degrees of freedom, we prepared a registered report in which all sample, design, procedure and analysis elements were decided prior to conducting the study.

Still, while the common impression in the literature is that the suppression-induced forgetting effect in the TNT task is a robust phenomenon in the healthy adult population (see Anderson & Huddleston, 2012; Stramaccia et al., 2021; and the preprint of Marsh & Anderson, 2022 for a comprehensive overview), we found null results with effect sizes close to zero (as opposed to average effect sizes of 0.36 in both the same-probe and independent-probe tests in a non-published meta-analysis by Clark, 2021). Therefore, in the following, we will discuss first, whether procedural elements of our replication may have led to the observed null results; second, whether the literature overestimated the effect size and robustness of suppression-induced forgetting in the TNT; and lastly, whether there are significant differences between previous TNT studies showing evidence of suppression-induced forgetting and our replication and how future research should tackle these issues to further examine suppression-induced forgetting in the TNT task.

To our knowledge, the current study is the first to apply the TNT task to an *online experimenter-present* setting. Even though this setting aided participant availability and eased data collection during the COVID-19 pandemic, it also introduced new challenges. The online procedure required attention to time differences, participant equipment and setup, and potential cultural barriers in communication and organisation. During the session, to minimise distraction and to ensure that participants did not disengage from the task, the experimenter guided the participants

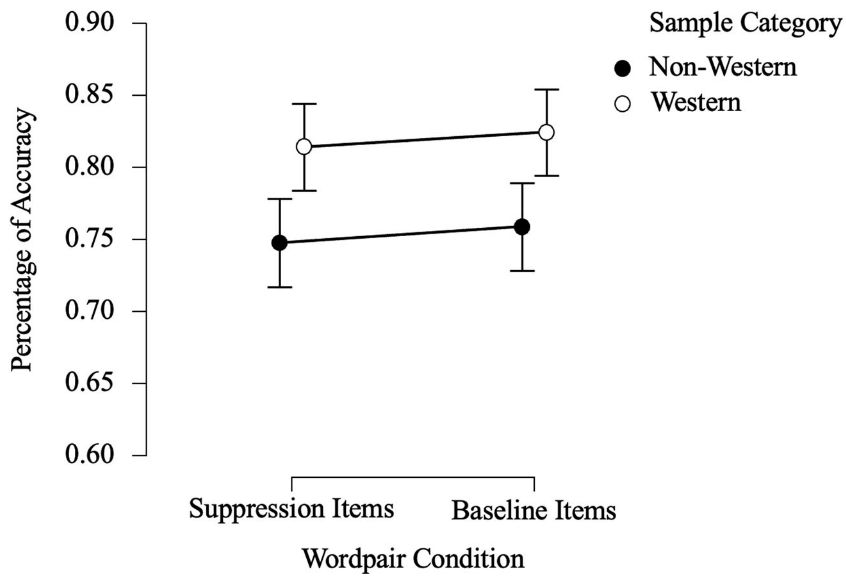

Figure 3. Differences between the same-probe suppression condition average accuracy and baseline condition average accuracy divided by the sample category with errors bars representing the 95% confidence interval.

throughout the online procedure while watching the participants' faces (front-on from the laptop/computer camera) and the home background. In doing so, participants may have felt "watched". In turn, this feeling may have increased the discomfort felt during the session. Yet, it is important to note that participants successfully adhered to the TNT phase instructions ($M = 0.97$, $SD = 0.03$; range = 0–1; higher numbers indicating more TNT instruction adherence) and evaluated the session favourably ($M = 9.95$, $SD = 0.81$; range = 9–27; lower numbers indicating a more favourable session evaluation).

Another potential difference between our replication and the prior literature reflects the high exclusion rate during the learning phase (39 participants, 18.31%) in our study. The TNT literature does not consistently report learning phase exclusions, or if they do, these exclusion rates seem to be notably lower (e.g., Bulevich et al., 2006; Hertel & Calcaterra, 2005; Wessel et al., 2020). Excluding participants who failed to learn the word pairs in time is essential for the procedure but might still induce a *selection bias* (i.e., systematic error due to non-random sampling from a population, e.g., Ellenberg, 1994) to the suppression-

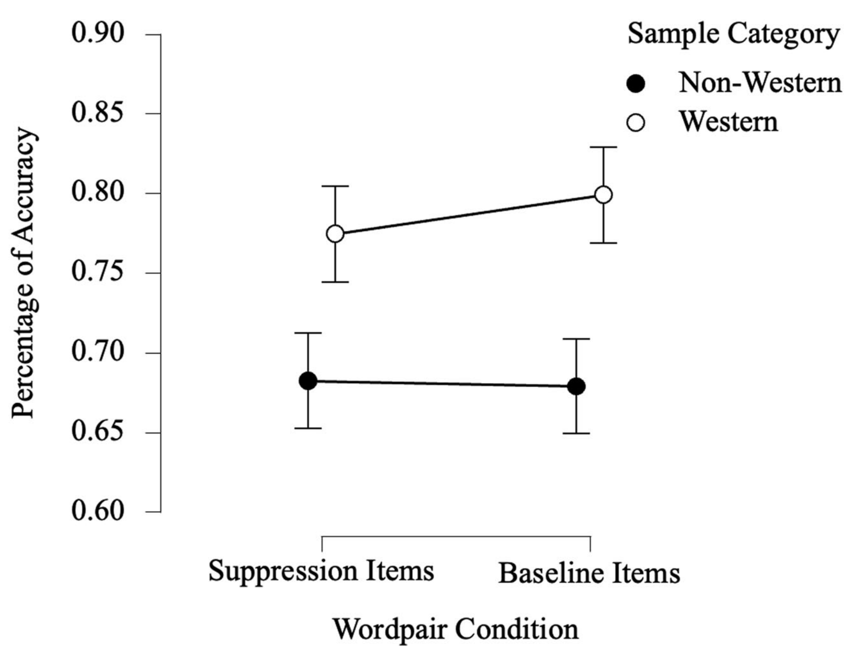

Figure 4. Differences between the independent-probe suppression condition average accuracy and baseline condition average accuracy divided by the sample category with error bars representing the 95% confidence interval.

Table 2. Means, standard deviations and Pearson's correlations with confidence intervals and Bayes factors.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. Compliance Score	1.32	1.40									
2. Same-Probe Difference Score	0.01	0.19	<.01 [−.16, .16] BF_{10} = 0.10								
3. Independent-Probe Difference Score	0.01	0.18	.01 [−.15, .17] BF_{10} = 0.10	.12 [−.05, .27] BF_{10} = 0.27							
4. TNT Phase Accuracy	0.97	0.03	.07 [−.09, .23] BF_{10} = 0.14	−.18 [−.33, −.02] BF_{10} = 1.17	.06 [−.10, .21] BF_{10} = 0.13						
5. Test Feedback Trial Count	124.23	46.74	−.05 [−.21, .11] BF_{10} = 0.13	.08 [−.08, .24] BF_{10} = 0.17	.01 [−.15, .17] BF_{10} = 0.10	−.55* [−.65, −.43] BF_{10} = 3.45e+10					
6. Substitution Strategy Use	3.53	1.49	−.19 [−.34, −.03] BF_{10} = 1.54	−.01 [−.17, .15] BF_{10} = 0.10	.09 [−.07, .25] BF_{10} = 0.20	−.12 [−.27, .05] BF_{10} = 0.27	.07 [−.09, .23] BF_{10} = 0.15				
7. Executive Functions Problems	9.31	3.25	.12 [−.04, .28] BF_{10} = 0.31	.03 [−.13, .20] BF_{10} = 0.11	−.05 [−.21, .11] BF_{10} = 0.13	.13 [−.04, .28] BF_{10} = 0.33	−.09 [−.25, .07] BF_{10} = 0.19	−.17 [−.30, .01] BF_{10} = 0.85			
8. Perceived Stress	22.35	8.36	.09 [−.07, .24] BF_{10} = 0.18	.02 [−.15, .18] BF_{10} = 0.10	−.11 [−.26, .06] BF_{10} = 0.23	−.08 [−.23, .08] BF_{10} = 0.16	.01 [−.15, .17] BF_{10} = 0.10	−.04 [−.20, .12] BF_{10} = 0.12	.44* [.30, .56] BF_{10} = 400536.78		
9. Thought Control Ability	83.10	16.86	−.05 [−.21, .11] BF_{10} = 0.13	−.03 [−.19, .13] BF_{10} = 0.11	.18 [.02, .33] BF_{10} = 1.03	−.07 [−.22, .10] BF_{10} = 0.14	.06 [−.10, .22] BF_{10} = 0.13	.15 [−.02, .30] BF_{10} = 0.48	−.42* [−.55, −.27] BF_{10} = 87289.91	−.72* [−.79, −.64] BF_{10} = 2.25e+10	
10. Session Evaluation	9.95	0.81	.07 [−.09, .23] BF_{10} = 0.15	.08 [−.08, .24] BF_{10} = 0.16	.05 [−.11, .21] BF_{10} = 0.13	−.14 [−.29, .02] BF_{10} = 0.40	−.01 [−.17, .15] BF_{10} = 0.10	−.14 [−.29, .02] BF_{10} = 0.40	.23* [.07, .38] BF_{10} = 5.77	.21 [.05, .36] BF_{10} = 2.48	−.23* [−.38, −.07] BF_{10} = 5.29

Note: *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. * indicates conclusive Bayesian evidence for the alternative hypothesis over the null hypothesis, determined by our pre-specified cut-off of 5. For all bivariate correlations with executive functions problems, three participants that failed the attention check, were excluded.

induced forgetting literature using the TNT task. Possible explanations for high number exclusion rates in the learning phase could be that (a) the word pairs were difficult to memorise, (b) the learning phase included too many word pairs, or (c) the instructions were unclear. However, these issues seem unlikely as the word pair stimuli set and the session instructions were retrieved and adapted from previous research that successfully replicated the same-probe effect (originally from Anderson & Green, 2001; Benoit & Anderson, 2012; adapted from Wessel et al., 2020).

An alternative account for the observed null result in the current study is that prior studies might have overestimated the effect size and the robustness of suppression-induced forgetting in the TNT. Meta-analysis results are important in statistically aggregating quantitative summaries of the literature, but their interpretation may be clouded if the literature is biased. Therefore, pre-registered replications are essential and can complement the knowledge derived from meta-analyses by providing unbiased effect size estimates. Even though both approaches aim to estimate the true effect size of a phenomenon, recent work shows that they tend to be discrepant (Kvarven et al., 2020; Lewis et al., 2022; Nieuwenstein et al., 2015). According to this prior research, the meta-analytic average effect size estimates tend to be, on average, three times larger than their replication counterparts. It is argued that this does not discount either approach and can be explained, at least to some extent, by publication bias, increased heterogeneity across studies, and post-hoc data-driven decision-making in the statistical analyses (Gelman & Loken, 2014). For example, the multiverse analysis by Wessel et al. (2020) suggested that the interpretation of TNT results may change depending on the different statistical methods employed. Nonetheless, if our null findings critically depend on our pre-registered data-analytic decisions, this would still call into question the robustness of suppression-induced forgetting in the TNT.

From the above considerations, the sole obvious difference between previous research and our replication attempt is the online application of the procedure. Therefore, we cannot rule out the possibility that this difference drives the observed null result. Although there is no reason to assume reduced compliance, motivation and, thereby, reduced suppression performance by the setup alone, a laboratory vs web comparison of the current procedure might be beneficial to eliminate this explanation of results. Further, we note that, initially, we chose the direct suppression instruction (as opposed to unaided suppression or thought substitution; e.g., Anderson & Green, 2001; Hertel & Calcaterra, 2005, respectively) because, theoretically, only this instruction should induce suppression-induced forgetting in both tests (Anderson & Levy, 2007; Bergström et al., 2009). However, the literature suggests that the suppression-induced forgetting effect in the same-probe test might be stronger in aided thought substitution strategies compared to direct suppression strategies (e.g., Hertel & Calcaterra, 2005). Therefore, we argue that it may be of interest to investigate multiple TNT operationalisations, including the effect of instructions.

Conclusion

The TNT task has been used widely in the laboratory to study suppression-induced forgetting. Yet, mixed results have been reported in the literature. To contribute to the accumulating scientific knowledge and facilitate more reliable estimates of suppression-induced forgetting for both the same- and independent-probe tests, we conducted a well-powered, pre-registered online experimenter-present version of the TNT task in English-speaking healthy individuals using a direct suppression instruction. We did not find evidence of suppression-induced forgetting in either test. Importantly, we acknowledge how our design choice, specifically the online nature of the procedure, could have influenced suppression-induced forgetting results. Nonetheless, considering the careful setup of the current study, including our quality and manipulation checks, our results may question the robustness of suppression-induced forgetting as assessed by the TNT paradigm.

Note

1. We conducted two pilot studies ($N = 15$; $N = 16$) before registering this report to verify the study procedure, estimate the study duration, ensure that the program works accurately, and data are correctly stored. The data of both pilot studies can be found on https://osf.io/e75a6/.

Acknowledgements

We thank Lilian Peetz and Lara Schönborn for their help in the data collection process for the pilot. Further, we thank Sophie Limmer, Aaron Lob, David Coba, Michal Broday and Robyn Lehmann for their help in the data collection process for the main study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Declarations of Interest: None

ORCID

Sera Wiechert http://orcid.org/0000-0002-0260-1702
Leonie Loewy http://orcid.org/0000-0002-4644-178X
Ineke Wessel http://orcid.org/0000-0001-6312-376X

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369. https://doi.org/10.1038/35066572

Anderson, M. C., & Huddleston, E. (2012). Towards a cognitive and neurobiological model of motivated forgetting. In R. F. Belli (Ed.), *True and false recovered memories: Toward a reconciliation of the debate* (pp. 53–120). Springer Science & Business Media.

Anderson, M. C., & Levy, B. J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In D. S. Gorfein, & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 81–102). American Psychological Association.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., Glover, G. H., & Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, *303*(5655), 232–235. https://doi.org/10.1126/science.1089504

Benoit, R. G., & Anderson, M. C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron*, *76*(2), 450–460. https://doi.org/10.1016/j.neuron.2012.07.025

Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage*, *48*(4), 726–737. https://doi.org/10.1016/j.neuroimage.2009.06.051

Bulevich, J. B., Roediger, H. L., Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no-think paradigm. *Memory & Cognition*, *34*(8), 1569–1577. https://doi.org/10.3758/BF03195920

Clark, C. (2021). *Forgetting as a consequence of retrieval suppression: a meta-analytic review of the think/no-think paradigm* (Doctoral dissertation, Memorial University of Newfoundland).

Conway, M. A. (2001). Repression revisited. *Nature*, *410*(6826), 319–320. https://doi.org/10.1038/35066672

del Prete, F., Hanczakowski, M., Bajo, M. T., & Mazzoni, G. (2015). Inhibitory effects of thought substitution in the think/no-think task: Evidence from independent cues. *Memory (Hove, England)*, *23*(4), 507–517. https://doi.org/10.1080/09658211.2014.907429

Depue, B. E., Banich, M. T., & Curran, T. (2006). Suppression of emotional and nonemotional content in memory. *Psychological Science*, *17*(5), 441–447. https://doi.org/10.1111/j.1467-9280.2006.01725.x

Dodier, O., Gilet, A.-L., & Colombel, F. (2021). What do people really think of when they claim to believe in repressed memory? Methodological middle ground and applied issues. *Memory*, *30*(6), 744–752. https://doi.org/10.1080/09658211.2020.1868524

Ellenberg, J. H. (1994). Selection bias in observational and experimental studies. *Statistics in Medicine*, *13*(5-7), 557–567. https://doi.org/10.1002/sim.4780130518

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, *102*(6), 460–465. https://doi.org/10.1511/2014.111.460

Hendriks, B., van Meurs, F., & Usmany, N. (2021). The effects of lecturers' non-native accent strength in English on intelligibility and attitudinal evaluations by native and non-native English students. *Language Teaching Research*. https://doi.org/10.1177/1362168820983145

Hertel, P., & McDaniel, L. (2010). The suppressive power of positive thinking: Aiding suppression-induced forgetting in repressive coping. *Cognition and Emotion*, *24*(7), 1239–1249. https://doi.org/10.1080/02699930903172377

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review*, *12*(3), 484–489. https://doi.org/10.3758/BF03193792

Hertel, P. T., & Mahan, A. (2008). Depression-related differences in learning and forgetting responses to unrelated cues. *Acta Psychologica*, *127*(3), 636–644. https://doi.org/10.1016/j.actpsy.2007.11.004

Horowitz, M. J. (1976). *Stress response syndromes*. Jason Aronson, Inc.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*(1), 2–9. https://doi.org/10.7771/1932-6246.1167

Joormann, J., Hertel, P. T., Brozovich, F., & Gotlib, I. H. (2005). Remembering the good, forgetting the bad: Intentional forgetting of emotional material in depression. *Journal of Abnormal Psychology*, *114*(4), 640–648. https://doi.org/10.1037/0021-843X.114.4.640

Joormann, J., Hertel, P. T., LeMoult, J., & Gotlib, I. H. (2009). Training forgetting of negative material in depression. *Journal of Abnormal Psychology*, *118*(1), 34–43. https://doi.org/10.1037/a0013794

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434. https://doi.org/10.1038/s41562-019-0787-z

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levy, B. J., & Anderson, M. C. (2012). Purging of memories from conscious awareness tracked in the human brain. *Journal of Neuroscience*, *32*(47), 16785–16794. https://doi.org/10.1523/JNEUROSCI.2640-12.2012

Lewis, M., Mathur, M. B., VanderWeele, T. J., & Frank, M. C. (2022). The puzzling relationship between multi-laboratory replications and meta-analyses of the published literature. *R. Soc. Open Sci*, *9*(2), 211499. https://doi.org/10.1098/rsos.211499

Marsh, L. C., & Anderson, M. (2022, June 27). Inhibition as a cause of forgetting. https://doi.org/10.31234/osf.io/6bwc2.

Meier, B., König, A., Parak, S., & Henke, K. (2011). Suppressed, but not forgotten. *Swiss Journal of Psychology*, *70*(1), 5–11. https://doi.org/10.1024/1421-0185/a000033

Murray, B. D., Muscatell, K. A., & Kensinger, E. A. (2011). Effects of emotion and age on performance during a think/no-think memory task. *Psychology and Aging*, *26*(4), 940–955. https://doi.org/10.1037/a0023214

Nieuwenstein, M. R., Wierenga, T., Morey, R. D., Wicherts, J. M., Blom, T. N., Wagenmakers, E. J., & van Rijn, H. (2015). On making the right choice: A meta-analysis and large-scale replication attempt of the unconscious thought advantage. *Judgment and Decision Making*, *10*(1), 1–17. https://doi.org/10.1017/S1930297500003144

Noreen, S., & Macleod, M. D. (2013). It's all in the detail: Intentional forgetting of autobiographical memories using the autobiographical think/no-think task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 375–393. https://doi.org/10.1037/a0028888

Noreen, S., & MacLeod, M. D. (2014). To think or not to think, that is the question: Individual differences in suppression and rebound effects in autobiographical memory. *Acta Psychologica*, *145*, 84–97. https://doi.org/10.1016/j.actpsy.2013.10.011

Otgaar, H., Howe, M. L., Dodier, O., Lilienfeld, S. O., Loftus, E. F., Lynn, S. J., Merckelbach, H., & Patihis, L. (2021). Belief in unconscious repressed memory persists. *Perspectives on Psychological Science*, *16*(2), 454–460. https://doi.org/10.1177/1745691621990628

Otgaar, H., Howe, M. L., Patihis, L., Merckelbach, H., Lynn, S. J., Lilienfeld, S. O., & Loftus, E. F. (2019). The return of the repressed: The persistent and problematic claims of long-forgotten trauma. *Perspectives on Psychological Science*, *14*(6), 1072–1095. https://doi.org/10.1177/1745691619862306

Patihis, L., Ho, L. Y., Tingen, I. W., Lilienfeld, S. O., & Loftus, E. F. (2014). Are the "memory wars" over? A scientist-practitioner gap in beliefs about repressed memory. *Psychological Science*, *25*(2), 519–530. https://doi.org/10.1177/0956797613510718

Paz-Alonso, P. M., Ghetti, S., Matlen, B. J., Anderson, M. C., & Bunge, S. A. (2009). Memory suppression is an active process that improves over childhood. *Frontiers in Human Neuroscience*, *3*(24), 1–6. https://doi.org/10.3389/neuro.09.024.2009

Racsmány, M., Conway, M. A., Keresztes, A., & Krajcsi, A. (2012). Inhibition and interference in the think/no-think task. *Memory & Cognition*, *40*(2), 168–176. https://doi.org/10.3758/s13421-011-0144-6

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

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Stramaccia, D. F., Meyer, A.-K., Rischer, K. M., Fawcett, J. M., & Benoit, R. G. (2021). Memory suppression and its deficiency in psychological disorders: A focused meta-analysis. *Journal of Experimental Psychology: General*, *150*(5), 828–850. https://doi.org/10.1037/xge0000971

Taubenfeld, A., Anderson, M. C., & Levy, D. A. (2019). The impact of retrieval suppression on conceptual implicit memory. *Memory (Hove, England)*, *27*(5), 686–697. https://doi.org/10.1080/09658211.2018.1554079

Van Elk, M., Matzke, D., Gronau, Q., Guang, M., Vandekerckhove, J., & Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: A skeptical perspective on religious priming. *Frontiers in Psychology*, *6*, 1365. https://doi.org/10.3389/fpsyg.2015.01365

van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory (Hove, England)*, *25*(9), 1201–1216. https://doi.org/10.1080/09658211.2017.1282518

van Schie, K., Geraerts, E., & Anderson, M. C. (2013). Emotional and non-emotional memories are suppressible under direct suppression instructions. *Cognition & Emotion*, *27*(6), 1122–1131. https://doi.org/10.1080/02699931.2013.765387

Vohs, K. D., Schmeichel, B. J., Gronau, Q. F., Finley, A., Wagenmakers, E., & Albarracín, D. (2021). A multi-site preregistered paradigmatic test of the ego depletion effect. *Psychological Science*, *32*(10), 1566–1581. https://doi.org/10.1177/0956797621989733

Waldhauser, G. T., Dahl, M. J., Ruf-Leuschner, M., Müller-Bamouh, V., Schauer, M., Axmacher, N., Elbert, T., & Hanslmayr, S. (2018). The neural dynamics of deficient memory control in heavily traumatized refugees. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-31400-x

Wessel, I., Albers, C. J., Zandstra, A. R. E., & Heininga, V. E. (2020). A multiverse analysis of early attempts to replicate memory suppression with the think/no-think task. *Memory (Hove, England)*, *28*(7), 870–887. https://doi.org/10.1080/09658211.2020.1797095

Williams, J. M. G., Stiles, W. B., & Shapiro, D. A. (1999). Cognitive mechanisms in the avoidance of painful and dangerous thoughts: Elaborating the assimilation model. *Cognitive Therapy and Research*, *23*(3), 285–306. https://doi.org/10.1023/A:1018743615228

Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., & Zeelenberg, R. (2018). Participant nonnaiveté and the reproducibility of cognitive psychology. *Psychonomic Bulletin & Review*, *25*(5), 1968–1972. https://doi.org/10.3758/s13423-017-1348-y

Appendix

Table A1. Key elements of the paradigmatic replication approach

Goals	Strategies	Rationale
Study formulation		
Public statement of intent	Registered report including hypotheses, methods, exclusion criteria and specified conclusions given different possible results	Open Science Practices Minimise researcher degrees of freedom
Choose implementation	Experimenter-present online	Especially during the SARS-COV2 pandemic, virtual data collection is valuable. Less laboratory use, more participant availability Experimenter makes sure virtually that participants engage and have an optimal study environment
Choose replication focus	Direct Suppression Instruction	Direct suppression (vs thought substitution) is more theoretically in line with both the same-probe and the independent-probe test
Choose sample	Fluent English speakers from countries in which English is (one of the) national languages	Generalizability across Western and non-Western countries (based on nationality), while holding experimental control over language proficiency and language of materials
Study and analysis preparation		
Methods testing and practice	Write and revise script for experimenters to follow	Reduce variation in procedural execution
Methods analysis	Write and test analysis script on pilot data	Reduce variation in analytical execution
Post data collection stage		
Ensure data integrity	Labs send data to one data handler Data handler merges and anonymises dataset	Ensure data integrity and increase confidence in results
Increase information value of data	After confirmatory data analysis, collaborating authors can suggest exploratory analyses	Perform tests unspecified in the registered report