Memory & Cognition (2025) 53:2028–2040
https://doi.org/10.3758/s13421-025-01696-y

Slowdown vs. breakdown of memory recall by retrieval stopping

Satoru Nishiyama[1,2] · **Randall C. O'Reilly**[3,4,5] · **Satoru Saito**[1]

Accepted: 7 February 2025 / Published online: 28 February 2025
© The Psychonomic Society, Inc. 2025

Abstract
Think/No-Think studies have shown that people can prevent memories from coming up to their mind by explicitly attempting to not think of them. However, there is an important limitation in the measures typically used: binary recall vs. no-recall accuracy under a specific time deadline. In this study, we instead focused on recall latency with a longer response window to accommodate a wider range of recall latencies. We found in Experiment 1 that direct suppression in the standard No-Think condition had a relatively uniform, graded effect, slowing the recall process in such a way that more recall failures occur with a short deadline, but a longer deadline (10 s) allows for successful recall at rates comparable to a baseline condition. In Experiment 2, thought substitution also caused the slowdown of the recall despite still lower recall rate than a baseline in 10 s. These results suggest that memory recall is subject to graded impairment across all items in a consistent manner instead of the breakdown of recall among a subset of memories. For understanding forgetting by retrieval stopping, excessive use of recall rate should be avoided, and recall latency is a potential alternative.

Keywords Retrieval stopping · Think/No-Think · Recall latency · Inhibition · Interference

It is widely believed that people tend to suppress unpleasant or otherwise unwanted memories, and considerable experimental support for such a process has been published (Anderson & Hulbert, 2021). For example, studies with the Think/No-Think (TNT) paradigm have shown that explicit attempts to not think of memories (hereinafter, *retrieval stopping*) can induce memory inhibition, resulting in forgetting as well as emotional and perceptual degradation of memories (Anderson & Green, 2001; Gagnepain et al., 2014, 2017; Kim & Yi, 2013; Nishiyama & Saito, 2022). The TNT literature has also demonstrated possible neural mechanisms of retrieval stopping (Anderson & Hanslmayr, 2014) and associated deficits in clinical disorders (e.g., Mary et al., 2020; Stramaccia et al., 2021). However, it is still unclear precisely what impact the retrieval stopping process has on memory recall. Does it break down (strongly disrupt) the recall of target memories, which results in long-lasting inaccessibility of the memories? Alternatively, are there perhaps more graded effects that might be suggestive of something, such as the slowdown of target recall, other than the breakdown of memory recall? Furthermore, it is typically assumed that items that are still recalled despite the effects of retrieval stopping somehow evaded the recall breakdown. However, what if instead *all* of the target memories were consistently weakened in a more graded way? More broadly, what kind of changes in memory/recall processes are referred to as "forgetting" in TNT studies?

The existing TNT studies report significant effects showing outright failures to recall some of the memories subject to the retrieval stopping, relative to baseline items. However, it is possible that the experimental settings in the recall phase of these studies could contribute to such findings—in particular, the use of a fixed deadline within which a response must be made. If the memories subject to retrieval stopping are consistently degraded in some way, then they

✉ Satoru Nishiyama
nishiyama.satoru.73z@kyoto-u.jp

✉ Satoru Saito
saito.satoru.2z@kyoto-u.ac.jp

[1] Graduate School of Education, Kyoto University, Sakyo-Ku, Kyoto, Yoshidahonmachi 606-8501, Japan

[2] Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-Ku, Tokyo 102-0083, Japan

[3] Department of Psychology, University of California, 135 Young Hall, One Shields Avenue, Davis, CA 95616, USA

[4] Center for Neuroscience, University of California, Grey Hall, 1544 Newton Court, Davis, CA 95618, USA

[5] Department of Computer Science, University of California, One Shields Avenue, Davis, CA 95616, USA

may take longer to recall and thus are differentially affected by this response deadline, with random variability determining which items are recalled within the deadline. This is the question we address in the two experiments reported here, which employed two different widely-used TNT paradigms, but with a significantly longer response deadline (10 s instead of the 4 s typically used).

Typical TNT experiments consist of three phases. First, participants study pairs of memory items, one of which is called the *cue* and the other is the *target*. Then, in the TNT phase, participants are presented with cues and required to recall the corresponding targets (Think condition) or to stop thinking about the targets (No-Think condition). Finally, participants perform recall tests where they attempt to recall the targets from the cues presented on the display, regardless of the instructions in the TNT phase. The benchmark finding of the TNT procedure is below-baseline forgetting in the final cued-recall test, which refers to lower recall rates of No-Think targets than that of the targets that are studied in the first study phase but neither recalled nor avoided in the TNT phase (Baseline targets). To obtain the recall rate, recall performance of each target is labeled as one of two classes: recall success or failure within a fixed response deadline, which is typically 4 s (Nardo & Anderson, 2024), but some studies set it to around 3 s (Benoit & Anderson, 2012; Nemeth et al., 2014; van Schie et al., 2013) or 5 s (Anderson et al., 2004, 2011; Bergström et al., 2009; Berkman et al., 2009). Although the recall rate (or number of recalled items) has been useful in a broad range of memory studies, this binary classification does not sufficiently capture the dynamic nature and operating principles of memory retrieval involved in recall processes (i.e., memory activation), which in turn limits the understanding of mechanisms underlying memory impairment by retrieval stopping.

Consider the recall latency underlying the recall rates shown in the bar plot in Fig. 1A, which is a typical result from TNT studies. As illustrated, there are two qualitatively different underlying distributions of recall over time that could give rise to the same pattern of results at deadlines around the standard 4 s. In one case, there is an overall reduction in the area under the curve for the No-Think targets along with a relatively small shift in the RT distribution to account for some slowing of these items (Fig. 1B). That is, recall of a subset of No-Think targets is broken down.[1] However, the same pattern of recall around the 4-s deadline can be produced by just a pure slowing of the No-Think distribution, with no change in total area under the curve (Fig. 1C). In this case, there is no residual difference in accuracy at the longer deadline (e.g., 10 s; Fig. 1E), in contrast to the case of the recall breakdown, where the difference in total area under the curve causes a persistent reduction in No-Think recall (Fig. 1D). In the pure slowing case, it is clear that retrieval stopping produces a consistent overall effect on successfully recalled and forgotten items, as contrasted with the partial impairment of a subset of items. Thus, this figure illustrates the importance of measuring the full distribution of the recall curves in order to obtain a better understanding of the nature of the underlying effects of retrieval stopping.

Investigating the distributions of recall latency could reveal different characteristics of memory impairment for the two types of No-Think strategies: direct suppression and thought substitution. Direct suppression refers to a No-Think strategy without any help from diversionary thoughts or other forms of interference (e.g., Bergström et al., 2009; Catarino et al., 2015). On the other hand, thought substitution refers to retrieval stopping with diversionary thoughts about something other than the target (e.g., del Prete et al., 2015; Hertel & Calcaterra, 2005). Both strategies result in below-baseline forgetting to the same extent in non-clinical samples (Stramaccia et al., 2021), although different mechanisms are involved (Benoit & Anderson, 2012). Specifically, forgetting by direct suppression is attributed to memory inhibition (Anderson & Green, 2001; Anderson & Hulbert, 2021; Taubenfeld et al., 2019; Wang et al., 2019). Forgetting by thought substitution is attributed to interference from diversionary thoughts (Wang et al., 2015), although some studies reported that thought substitution also causes memory inhibition (Benoit & Anderson, 2012; del Prete et al., 2015). While we did not have any specific prediction on which breakdown or slowdown is associated with each of the strategies and underlying mechanisms (inhibition and interference), observed distributional patterns could demonstrate the similarities and differences of the strategies discussed above.

To determine which of the two alternative distributions shown in Fig. 1 is driving the observed below-baseline forgetting, we conducted two experiments using a long deadline (10 s) that should allow a more complete estimation of the full underlying distributions across conditions. As an important replication check, we are still able to simulate recall rates at deadlines that have been employed in previous studies, such as 3, 4, and 5 s, to compare our results with prior studies. Although several previous studies measured recall latency (Cano & Knight, 2016; Haghighi et al., 2020; Nishiyama & Saito, 2021; Salamé & Danion, 2007), those studies did not examine the full distributions as we do here. Haghighi et al. (2020) pointed out the issue of trial duration in the recall tests and collected recall latency with long trial duration, but their focus was on ERP analyses of neural signals, not the distinctions we focus on here.

[1] The "breakdown" of recall is not necessarily derived from loss or removal of memory traces (i.e., strong from of forgetting by Davis et al., 2007).

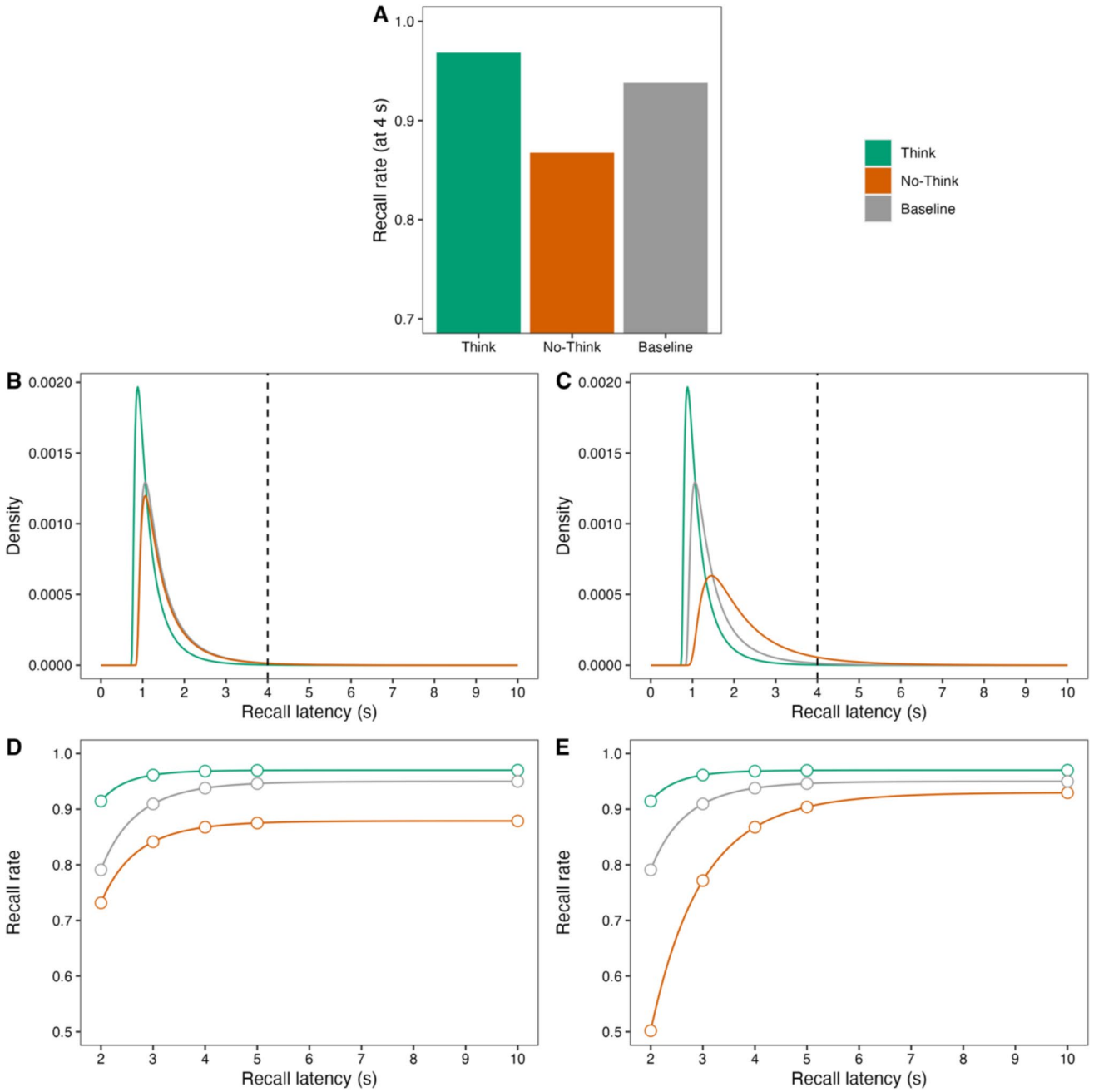

Fig. 1 Illustration of possible recall latency distributions that produce the typical patterns of recall rates in the TNT paradigm. **A** Typical recall rates for the TNT paradigm, which are typically obtained in recall tests with a duration of 4 s. **B** A possible distribution of recall latencies that produce the typical recall rates depicted in Panel A. The distribution for No-Think targets is characterized by an overall reduction in the area under the curve along with a relatively small shift in the RT distribution, assuming the recall breakdown of a few No-Think targets. **C** Another possible distribution of recall latencies that produce the typical recall rates depicted in Panel A. The distribution for No-Think targets is characterized by the slowdown of recall with no change in total area under the curve, assuming overall slowdown of recall of No-Think targets. **D** Cumulative recall rate corresponding to the area under the curve of the latency distribution in the case of recall breakdown (as depicted in Panel B). The recall rate of the No-Think targets is lower than the other targets in 10 s as well as in 4 s. **E** Cumulative recall rate corresponding to the area under the curve of the latency distribution in the case of recall slowdown (as depicted in Panel C). Below-baseline forgetting is observed in 4 s, but it is not in 10 s. (Color figure online)

In the two experiments, different instructions were provided for the critical retrieval stopping manipulation in the No-Think condition to target the potential differences between direct suppression vs. thought substitution. Participants were given the direct suppression instruction in Experiment 1, and the thought substitution instruction in

Experiment 2, but the methods were otherwise identical, and the same as used in many other published TNT studies, except for the use of the longer 10-s response deadline.

Method

The current experiments were approved by the Institutional Ethics Committee for Experimental Psychology Research at the Graduate School of Education, Kyoto University (approval number: CPE-203). All participants provided written informed consent in accordance with the protocol. Sample sizes of 36 were determined by a priori power analysis with G*Power 3.1 (Faul et al., 2007) as well as by the number of counterbalancing conditions for the assignment of items to the conditions (three conditions) and for the order of final recall tests (two conditions). The power analysis ensured that the statistical power was higher than 80% using a two-sided paired *t* test with a significance level of 5% when the effect size of the difference between No-Think and Baseline was medium ($d_z = 0.5$) and the sample size was larger than 34.[2] As reported below, the number of participants tested was more than the target sample size (i.e., 36) in both experiments. The reason was that some participants were excluded based on some criteria (see below in detail), and additional participants were recruited to achieve the target sample size. The exclusion criteria were set before data collection, and the recruitment procedure was done before data analysis.

Participants

Experiment 1 Forty-three university students participated in Experiment 1. They were compensated with a book coupon equivalent to 1,000 Japanese yen. All participants were native Japanese speakers with no history of mental disorders. Data from seven participants were excluded from analyses for the following reasons: The vocal responses of five participants were not collected in some trials of test phases, and two participants did not correctly answer the questions about the TNT instructions at the end of the experiment. Thus, the final sample size was 36 (mean age = 20.53 years, *SD* = 1.87, men = 24).[3]

Experiment 2 Forty-six university students participated in Experiment 2, receiving a book coupon equivalent to 1,000 Japanese yen. All participants were native Japanese speakers with no history of mental disorders. Data from ten participants were excluded from analyses for the following reasons: Eight participants could not reach the criterion (80% correct recall) on the pre-TNT recall test, and other two participants did not correctly answer the questions about the TNT instructions at the end of the experiment. The final sample size was 36 (mean age = 20.39 years, *SD* = 1.86, men = 16).

Materials

The same materials were used in both experiments. Thirty-six sets of four words (cue, target, substitute, and independent probe) were used. Twelve additional sets of words were used as the fillers. All were emotionally neutral, highly familiar nouns (over 5.5 of 7 in a Japanese corpus; Amano & Kondo, 1999) with two to five morae. Words for the independent probe (IP) were category names of the corresponding targets. The target words were composed of Japanese letters (i.e., hiragana or katakana). Other kinds of words (cues, substitutes, and IPs) were composed of any types of letter forms, including Japanese characters (i.e., kanji), or combinations of them. Prior to the experiment, the association strength of each word pair in the word sets was evaluated by 19 university students who were naïve to the experimental purpose and did not participate in either of the two experiments (mean age = 21.58 years, *SD* = 2.43, men = 14). The participants rated the association strength of each word pair presented in a random order on the PC display on a 5-point scale (1 = *very weak* to 5 = *very strong*). Cues were weakly associated with targets and IPs (cue–target, mean = 1.41, *SD* = 0.39; cue–IP, mean = 1.55, *SD* = 0.59), and target-substitute associations were also weak (mean = 1.28, *SD* = 0.24). In contrast, cues were strongly associated with substitutes (mean = 4.35, *SD* = 0.41). Twelve word sets were assigned to each of the three TNT conditions (Think, No-Think, Baseline). The assignment was counterbalanced across participants using a Latin square design.

Procedure

Experiment 1 The experiment consisted of five phases: encoding, pre-TNT recall test, TNT, final recall test, and postexperimental questionnaire (Fig. 2). All phases, except the postexperimental questionnaire, were administered by PsychoPy (Version 1.85.3; Peirce, 2007). In the encoding phase, participants were presented with all 48 cue–target pairs for 5 s each and practiced recalling targets from cues. The presentation order of items was randomized, except that six items at the start and end of the sequence were always filler pairs. The participants then strengthened the associations through test and feedback cycles. In each trial of the cycle, participants were given a cue and were required to recall the corresponding target aloud

[2] d_z represents the standardized mean difference effect size for within-subjects designs (Lakens, 2013).

[3] We recruited additional participants to compensate the exclusion before the data analysis.

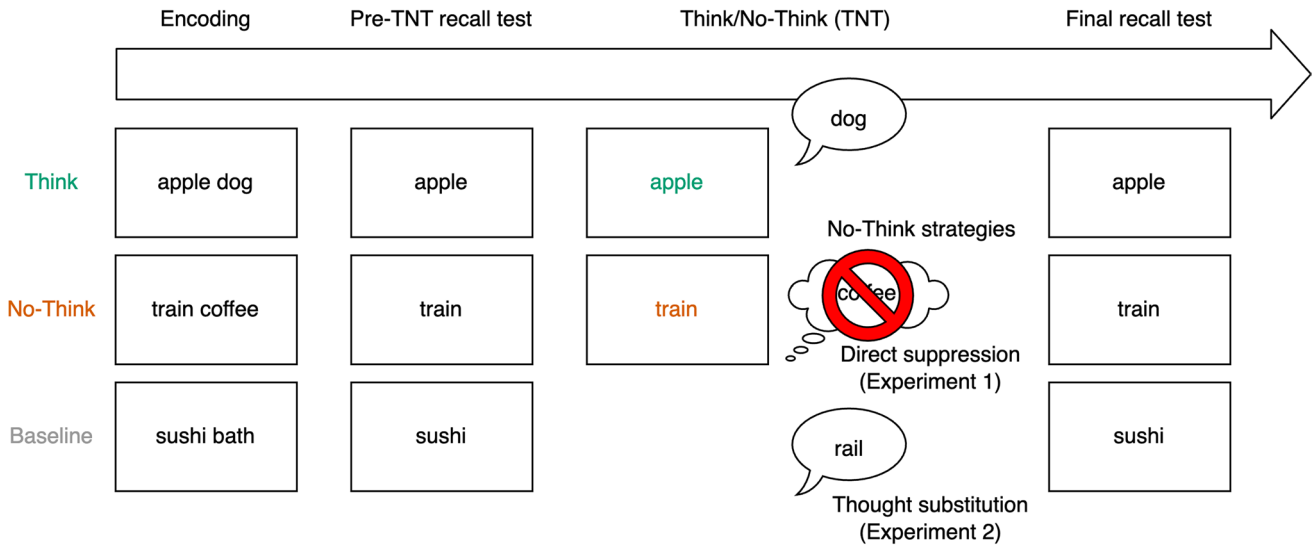

Fig. 2 Outline of the experimental procedure. The experiments were composed of five phases (the fifth phase, postexperimental questionnaire, is not shown in the figure). After participants studied word pairs (encoding phase), they performed a cued-recall test (pre-TNT recall test), in which they recalled aloud words (targets) paired with the presented words (cues). Then, participants performed the TNT task. They recalled aloud targets as in the cued-recall test when cues were presented in green (Think trials). Participants engaged in stopping retrieval of targets when cues were presented in red (No-Think trials). Different No-Think strategies were instructed in Experiments 1 and 2. In Experiment 1, participants attempted to stop retrieval of targets by pushing away any thoughts other than presented cues from their awareness (direct suppression). In Experiment 2, participants were instructed to recall substitutes, which were studied before the TNT task, instead of targets (thought substitution). Finally, participants recalled targets as in the pre-TNT recall test (final recall test). (Color figure online)

within 5 s. They pressed the space bar when they completed their vocal response. After participants pressed the space bar or 5 s elapsed, the correct target was presented as feedback regardless of whether their response was correct or not. Participants indicated whether their response was correct or not by pressing the corresponding keys, the ";" key, labelled with a circle (○) for correct, and the "a" key, labelled with a cross (×) for wrong. After they pressed either key, the feedback disappeared, and the next trial started. All cues were presented once during the first cycle. If participants indicated "wrong" for some targets, the corresponding cues were presented in the next cycle. This test and feedback cycle continued until they indicated "correct" for all pairs (mean cycles = 1.85, SD = 0.52). The presentation order of cues in the first cycle was randomized in the same way as in the initial encoding, and the presentation orders in the following cycles were fully randomized.

After the encoding phase, participants performed a cued-recall test without feedback (pre-TNT recall test). Participants were required to recall aloud targets associated with the cues presented on the display for 10 s. They pressed the space bar to continue to the next trial after they responded. All responses were recorded with Blue Microphones Snowball (Baltic Latvian Universal Electronics, LLC) and saved as separate audio files. Randomization of item presentation was administered in the same manner as the initial encoding (i.e., filler pairs were tested in the first and last six trials). In the pre-TNT recall test, participants had to recall at least 80% of the targets successfully to promote intrusion of targets during No-Think attempts (cf. Levy & Anderson, 2012). Otherwise, the data were excluded.

In the TNT task, participants engaged in retrieval (Think) or stopping retrieval (No-Think) of targets paired with cues on the display depending on the font color of the cues. For Think trials, cues were presented in green and participants were instructed to recall the corresponding targets aloud before the cues disappeared.[4] They continued to the next trial by pressing the space bar after they completed their vocal responses. For No-Think trials, cues were presented in red and participants were instructed to focus on the cues but prevent the corresponding targets from coming into their mind. It was emphasized that they should just focus on the cues and avoid thinking any other thoughts, including about the targets (direct suppression; cf. Bergström et al., 2009). After the instructions, participants were asked what they had to do when they saw green and red cues. If they did not remember the TNT instructions correctly, the instructions were repeated. Then, participants performed 12 practice

[4] Not all the TNT studies required participants to recall the targets aloud in the Think trials. We employed the current procedure in order to more likely keep participants' attention to the task than in the case where participants perform the TNT task just internally.

trials (six Think and six No-Think trials) with filler pairs. Each cue was presented once for 3 s. After the practice, participants performed the main 384 trials with the cue–target pairs assigned to the Think or No-Think conditions. The main trials were divided into 16 blocks. In each block, the 24 cues were each presented one at a time. The order of the Think and No-Think trials was randomized. Participants took a break every after they completed two blocks.

In the final recall test phase, participants performed two cued-recall tests: the same probe (SP) test and independent probe (IP) test. The SP test was conducted in the same way as in the pre-TNT recall test. Prior to the SP test, participants were instructed to recall one of the studied words associated with the presented cue. They were also told that each cue was one of the studied pair. The procedure of the IP test was identical to the SP test, except that the reminder for each target was its category name and first letter. Prior to the IP test, participants were instructed to recall one of the studied words corresponding to the reminder on the display. They were also informed about the reminder. The order of the two tests was counterbalanced across the participants.

At the end of the experiment, participants answered three questions on Google Forms. First, participants were asked an open question about their impressions of the experiment. They typed their answers in a text input area. Second, participants were asked about the TNT instructions. They were given two questions on how they performed when words were presented in green/red and they typed their answers in the input areas below each question. Finally, the participants indicated how well they concentrated during the experiment, choosing one of four options ("I concentrated on the tasks throughout the experiment and never felt sleepy"; "I did not concentrate on the tasks fully, although I did not feel sleepy"; "I sometimes felt sleepy, although I did not fall asleep"; "I sometimes fell asleep"). If participants chose the last option, their data would be excluded from the analysis; however, no data were excluded for this reason.[5]

Experiment 2 The procedure was identical to that of Experiment 1, except for the following changes. First, thought substitution was given as an instruction for the No-Think trials in the TNT task. Participants were required to recall aloud a substitute word associated with the cue presented on the display if the cue's font color was red. They continued to the next trial by pressing the space bar after they responded. To perform thought substitution, participants studied cue–substitute pairs three times: after the instruction, after the TNT practice, and after the eighth block of the main TNT task was completed.

Second, in the instruction of the SP test, it was emphasized that to-be-remembered words were those learned in the study phase performed at the beginning of the experiment.

Third, in the final recall test phase, participants performed a cued-recall test for the substitute words after the SP and IP tests were completed. In each trial of the cued-recall test, participants were given a No-Think cue for 10 s each and were required to recall the corresponding substitute word aloud. They pressed the space bar to continue to the next trial after they responded. Their vocal responses were recorded. Prior to the cued-recall test, participants were instructed to recall one of the words studied in the middle of the experiment (i.e., substitutes) corresponding to the cue on the display.

Data analyses

The recall latency data obtained from the recall tests were subjected to the analyses described below. Recall latency was computed from an audio file of each trial with the help of Chronset (Roux et al., 2017), which detects the onset of the initial vocal response.[6] If participants gave more than one response in a trial and the final one was correct (i.e., they corrected their response), we computed the recall latency of the final response manually with Audacity 2.1.1 (Audacity Team, 2015). We conducted all statistical analyses in R (Version 4.0.2; R Core Team, 2020) with the package *rstatix* (Version 0.6.0; Kassambara, 2020). For repeated-measures analyses of variance (ANOVAs), degrees of freedom were corrected using the Greenhouse–Geisser procedure in case of a violation of the assumption of sphericity. For follow-up multiple *t* tests, *p* values were adjusted using Holm's method. All datasets and analysis codes for the current study are available at the Open Science Framework repository (https://osf.io/fuwhk/).

Distribution of recall latency across participants and log-transformed recall latency We first drew density plots for correct recall latency across the participants to investigate which of the two recall patterns described in the Introduction was obtained (Fig. 1). Kernel density estimation was implemented by the "stat_density" function of *ggplot2* (Version 3.3.3; Wickham, 2016), an R package. Gaussian kernel was used. Bandwidths were determined by Silverman's rule of thumb, which the "stat_density" applies by default. All

[5] Participants were informed that there was no penalty for the choice.

[6] Chronset is not validated for Japanese in a systematic way. However, a previous study that conducted onset estimation of Japanese oral recall by Chronset and manual reported that estimates by Chronset reliably matched ones by manual onset detection ($R^2 = .92$; Nishiyama & Saito, 2021).

latency data of correct recall from all the participants were included to estimate a kernel density for each condition.

To ensure that the recall latency enhanced or impaired memory recall through the Think and No-Think attempts, mean log-transformed recall latency, which might represent the central tendency of the recall performance, was subjected to statistical analyses. The log-transformation was conducted to correct the skewness of the recall latency distribution.[7] Mean log-transformed recall latency was computed for each participant. We included the trials where participants recalled correct targets within 10 s.[8] For mean log-transformed latency of the pre-TNT recall test and SP test, we conducted a 3 (TNT treatment: Think, No-Think, and Baseline) × 2 (Test: pre-TNT and SP) repeated-measures ANOVA. We included the Test factor because the two tests with the same materials and procedures could be contrasted and because comparing the performance of the two tests (i.e., performance before and after the TNT phase) was expected to show the effect of memory control more clearly than conducting a statistical test only on the SP test. For log-transformed latency of the IP test, we conducted a one-way repeated-measures ANOVA, where TNT treatment served as a within-subject factor.

Recall rate as a function of elapsed time To describe the recall rates as a function of time, we computed recall rates at given time points (2, 3, 4, 5, and 10 s) from the recall latencies. If targets were not recalled successfully within a given time (i.e., wrong targets were recalled or the recall latency was longer than the time), they were labelled as forgotten. Otherwise, the targets were labelled as successfully recalled. To show that the observed size of forgetting depends on the trial duration of the recall tests, recall rates at 4 s and 10 s of the pre-TNT and SP tests were subjected to 3 (TNT treatment) × 2 (Test) repeated-measures ANOVA, respectively. Likewise, we conducted one-way repeated-measures ANOVAs with a within-factor (TNT treatment) for recall rates at 4 s and 10 s of the IP test. 4 s was selected as the counterpart of 10 s because 4 s has been the standard trial duration for recall tests in the TNT paradigm (Nardo & Anderson, 2024). The reason for the separate ANOVAs was that the recall rates at 4 s and 10 s were not independent of each other.

Results

Distribution of recall latency across participants and log-transformed recall latency

The density plots (Figs. 3A and 4A) describe the distributions of recall latency across all participants in Experiments 1 and 2, respectively. The distributions were right-skewed across tests, TNT manipulations, and No-Think instructions. Although the distributions of the three conditions overlapped before the TNT task (i.e., in the pre-TNT recall test), different distributions were clearly observed in the same probe (SP) test, which was performed after the TNT task with the same cues as used in the pre-TNT recall test and in the TNT phase. Think targets were recalled faster than Baseline targets. In contrast, No-Think targets were recalled more slowly than Baseline targets. Consistent with the shapes of the distributions, in the SP test, mean log-transformed recall latency, which represents the central tendency of the distribution, was longer in the No-Think condition than in the Think and Baseline conditions, and it was shorter in the Think condition than in the Baseline condition (Figs. 3B and 4B). For mean log-transformed recall latency in Experiment 1, a 3 (TNT treatment) × 2 (Test) ANOVA showed that the interaction was significant, $F_{1.71,59.94} = 37.11$, $p < 0.001$, $\eta_G^2 = 0.12$. A simple main effect analysis revealed that the main effect of TNT treatment was not significant for the pre-TNT recall test, $F_{2,70} = 0.86$, $p = 0.428$, $\eta_G^2 = 0.01$, whereas it was significant for the SP test, $F_{2,70} = 43.75$, $p < 0.001$, $\eta_G^2 = 0.28$. Post hoc pairwise t tests for the SP test performance demonstrated that the mean log-transformed recall latency was significantly longer in the No-Think condition than in the Think and Baseline conditions, Think vs. No-Think, $t(35) = 8.67$, $p < 0.001$, $d_z = 1.45$, 95% CI [0.97, 1.91]; Baseline vs. No-Think, $t(35) = 4.20$, $p < 0.001$, $d_z = 0.70$, 95% CI [0.33, 1.06]. Furthermore, the mean log-transformed recall latency for Think targets was shorter than that for Baseline targets, $t(35) = 5.95$, $p < 0.001$, $d_z = 0.99$, 95% CI [0.59, 1.39]. These results indicate that deliberate recall (i.e., Think attempts) facilitated recall, and direct suppression hindered memories from being recalled.

[7] Although there are many options to analyze RT data, we decided to apply to the RT data log transformation instead of distributional approaches, such as ex-Gaussian fitting, because the number of observations was small for each participant (at most 12 observations per condition). Previous studies showed that smaller observations make estimation performance worse (e.g., Lacouture & Cousineau, 2008). Moreover, a recent paper (Blanch & Cousineau, 2024) also justifies our analytic decision by indicating that at least 40 observations are required for ex-Gaussian fitting. For the same reason of the data size, we avoided computing mean RT after excluding outliers defined based on standard deviations or median absolute deviation.

[8] When calculating performance in the final recall test, there is an option to include only the targets that were recalled successfully in the pre-TNT test (conditionalization; Nardo & Anderson, 2024). The conditionalization helps to examine the effect of retrieval stopping on memory more clearly. However, it varies number and contents of stimuli across participants. The downside was considered more than the upside in the current study because we planned to contrast the results of pre-TNT and SP tests. For this purpose, the number of observations per cell (each condition for each participant) should be equal between the tests, especially for the analysis of recall rates. Thus, conditionalization was not conducted. Given that participants showed over 90% accuracy in the test, conditionalization would not change the results.

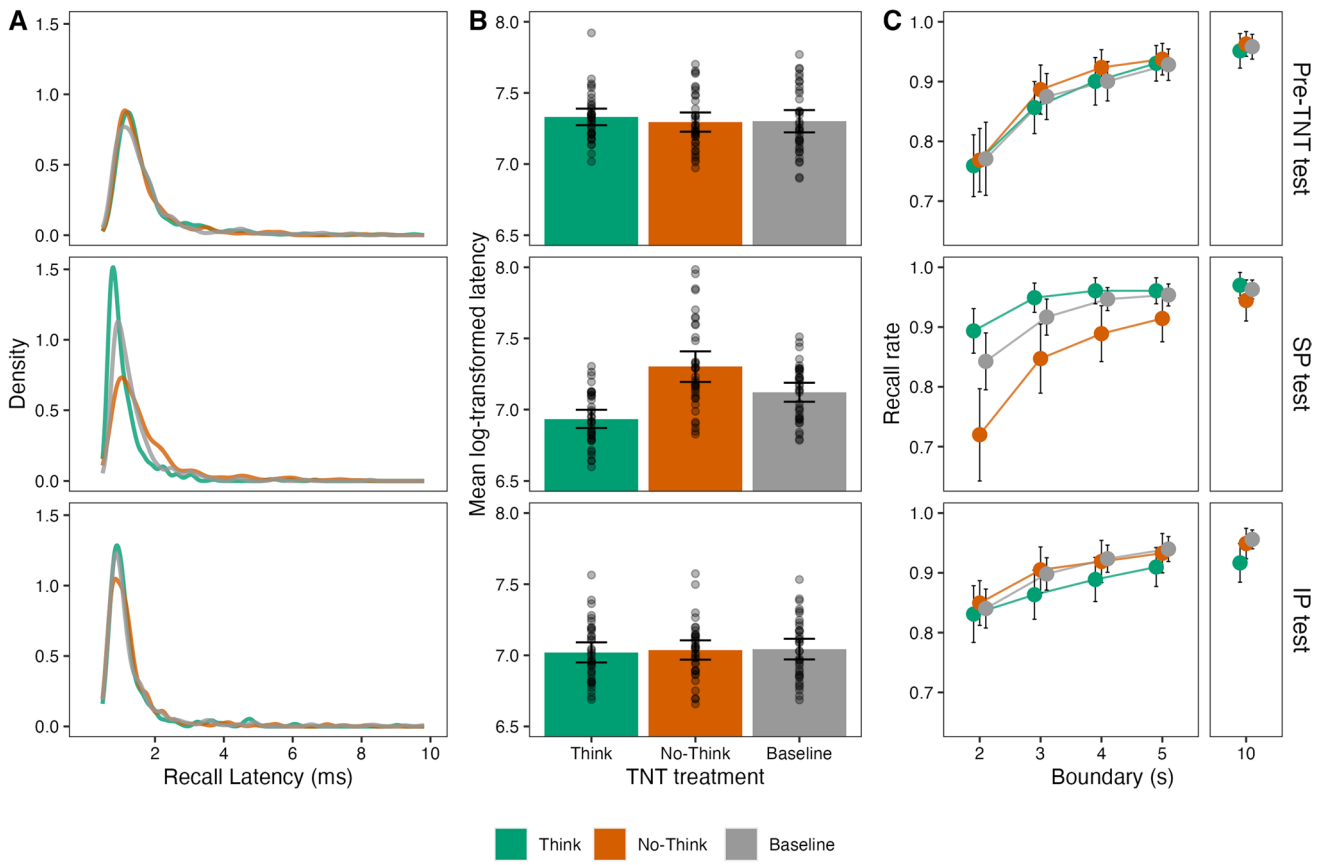

Fig. 3 Distributions of recall latency, mean log-transformed recall latency, and recall rates estimated from recall latency in Experiment 1. **A** Density plots of recall latency across participants ($N = 36$). **B** Bar plots of the mean log-transformed recall latency. The points represent individual data points. **C** Line graphs of cumulative recall rates at given time points. Error bars in Panels B and C show the 95% confidence intervals. SP test = same probe, IP test = independent probe. Note that No-Think effects were not found in the IP test (see discussion). (Color figure online)

Similarly, for mean log-transformed recall latency in Experiment 2, a 3 (TNT treatment) × 2 (Test) ANOVA showed that the interaction was significant, $F_{2,70} = 108.56$, $p < 0.001$, $\eta^2_G = 0.29$. A simple main effect analysis revealed that the main effect of TNT treatment was not significant for the pre-TNT recall test, $F_{2,70} = 0.40$, $p = 0.674$, $\eta^2_G = 0.00$, while it was significant for the SP test, $F_{2,70} = 154.56$, $p < 0.001$, $\eta^2_G = 0.61$. Post hoc pairwise t tests for the SP test performance demonstrated that the mean log-transformed recall latency was significantly longer in the No-Think condition than in the Think and Baseline conditions, Think vs. No-Think, $t(35) = 14.47$, $p < 0.001$, $d_z = 2.41$, 95% CI [1.76, 3.06]; Baseline vs. No-Think, $t(35) = 12.98$, $p < 0.001$, $d_z = 2.16$, 95% CI [1.56, 2.76]. Furthermore, the mean log-transformed recall latency for Think targets was shorter than for Baseline targets, $t(35) = 5.51$, $p < 0.001$, $d_z = 0.92$, 95% CI [0.52, 1.30]. Consistent with Experiment 1, Think attempts enhanced recall performance, and thought substitution hindered recall.

The distributions of recall latency in the IP test overlapped among the three conditions in both experiments (bottom row in Figs. 3 and 4). One-way ANOVAs for log-transformed recall latency showed that the main effect of TNT treatment was not significant in either experiment, Experiment 1, $F_{2,70} = 0.23$, $p = 0.795$, $\eta^2_G = 0.00$; Experiment 2, $F_{2,70} = 0.64$, $p = 0.530$, $\eta^2_G = 0.01$.

Recall rate as a function of elapsed time

We estimated the recall rates at each time point during the recall period. Figures 3C and 4C show that in both experiments the recall rates increased as a function of time, and reached almost 1 (100%) at a maximum of 10 s in all three conditions, regardless of the test timing (the pre-TNT or SP test). This clearly supports the slowdown account for memory impairment. In line with the slowdown, less below-baseline forgetting was observed at a later time point, such as 10 s. In Experiment 1, no reliable below-baseline forgetting

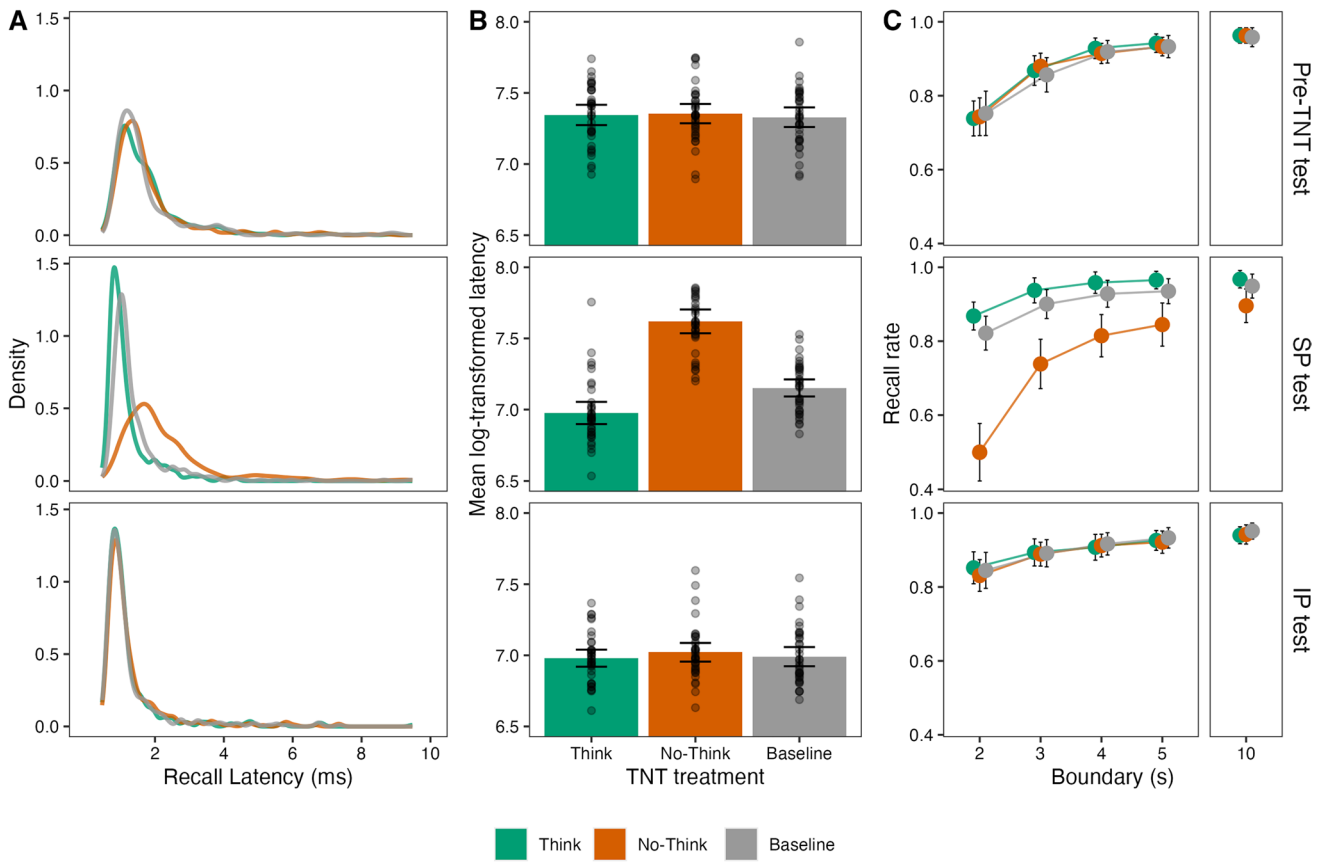

Fig. 4 Distributions of recall latency, mean log-transformed recall latency, and recall rates estimated from recall latency in Experiment 2. **A** Density plots of recall latency across participants ($N = 36$). **B** Bar plots of the mean log-transformed recall latency. The points represent individual data points. **C** Line graphs of cumulative recall rates at given time points. Error bars in the Panels B and C show the 95% confidence intervals. SP test = same probe, IP test = independent probe. As in Experiment 1, No-Think effects were not found in the IP test (see discussion). (Color figure online)

was observed at 10 s, $t(35) = 1.11$, $p = 0.588$, $d_z = 0.19$, 95% CI [−0.15, 0.51], whereas significant below-baseline forgetting was observed when the trial duration was as short as in the previous studies (e.g., at the standard test duration of 4 s), $t(35) = 2.38$, $p = 0.046$, $d_z = 0.40$, 95% CI [0.05, 0.73]. In Experiment 2, although there was significant below-baseline forgetting at 10 s, $t(35) = 2.60$, $p = 0.027$, $d_z = 0.43$, 95% CI [0.09, 0.77], the effect size was smaller than at 4 s ($d_z = 0.81$, 95% CI [0.43, 1.19]. These results suggest that the size of forgetting observed using a standard binary assessment method (i.e., recall rate) is influenced by the trial duration of the recall test.

In both experiments, recall rates in the IP test increased as a function of time, although there seemed to be no difference between the conditions (see the line plots in the bottom right corner of Figs. 3 and 4). One-way ANOVAs showed that the main effect of TNT treatment was not significant (F values < 0.3, p values > 0.7), except for recall rates at 10 s in Experiment 1, $F_{2,70} = 3.20$, $p = 0.047$, $\eta_G^2 = 0.05$. Post hoc t tests did not demonstrate any reliable difference between the conditions, after Holm's adjustments for the p values (t values < 2.1, p values > 0.13).

Discussion

The TNT literature has provided theoretical and clinical implications about the control of memory processes (Anderson & Hulbert, 2021). However, issues with recall rates have been understudied. Most TNT studies have tested the effect of memory control using recall rates, which evaluate the memory performance for each item using the two categories of success or failure, although memory recall is actually exhibited through nonbinary memory activation dynamics. Furthermore, the trial duration in recall tests, which potentially affects the recall rate, was arbitrarily determined in most studies. These issues must be considered in the theoretical investigation of retrieval stopping.

The primary goal of the two experiments was to determine the overall shapes of the recall distributions across

a long recall time window, in order to be able to distinguish between two qualitatively different effects of retrieval stopping in the TNT paradigm, as illustrated in Fig. 1. In both experiments, the distributions of recall latency were distinctive across the conditions of TNT manipulation in the same-probe (SP) test, which was performed after the TNT phase, in contrast to the overlapped distributions in the pre-TNT recall test. More Think targets were recalled faster, whereas more No-Think targets were recalled slower, compared with Baseline targets. The overall slower recall of No-Think targets suggests that forgetting by retrieval stopping results from the slowdown of memory recall of many targets, including the targets that were regarded as "recalled" in the conventional approach using recall rate, instead of the breakdown of memory recall of a few targets with other targets intact. In other words, retrieval stopping cannot induce long-lasting inaccessibility of unwanted memories, much less removal of them, but, at best, prevent them from coming up to the mind. This does not necessarily degrade the effectiveness of retrieval stopping. No-Think targets are recalled more slowly than Baseline targets even in the experimental settings where participants try to recall them. Moreover, people usually have little motivation to recall them for several seconds (e.g., 10 s) in everyday life (retrieval tendency hypothesis by Anderson & Huddleston, 2012). Thus, the slowdown effect of retrieval stopping can help us keep away from unwanted memories.

In line with the recall latency distributions, the results of the estimated recall rate demonstrated that more items were recalled as a function of elapsed time, and nearly all items were recalled at 10 s, regardless of manipulations during the TNT task (i.e., Think, No-Think, or Baseline). Greater recall rates in all conditions resulted in less below-baseline forgetting at 10 s despite reliable below-baseline forgetting at 4 s. Intriguingly, in Experiment 2, below-baseline forgetting was still observed at 10 s, although the differences were smaller than at 4 s. This may represent the breakdown of memory recall as illustrated by the distributions in Fig. 1B, suggesting that thought substitution impairs not only recall processes of No-Think targets in general, but also causes (nearly) complete loss of a few target memories due to strong interference from substitutes.

The results of recall latency indicate the similarity and difference between direct suppression and thought substitution. Both strategies slow down memory recall, and thought substitution breaks it down for a few targets. These might correspond to mechanisms underlying the two strategies that have been proposed in the literature. That is, the slowdown might be attributed to memory inhibition, which is assumed to underlie direct suppression and possibly thought substitution. Memories with inhibited representations are activated at lower level at the start of cue presentations in cued recall, resulting in more time to reactivate sufficiently for recalling them compared with Baseline targets. In contrast, the breakdown might be attributed to interference, which is assumed to underlie thought substitution. Spontaneous recall of nontarget memories interferes with that of target memories, resulting in longer latency. Associations between cues and substitutes formed by repeated substitution in the TNT phase would prompt recall of substitutes in response to cue presentations at the test phase.

The discussion above is based on the interpretation that lower recall rate of No-Think targets at 10 s in the thought substitution group was due to the breakdown of recall of a few forgotten targets. Alternatively, the below-baseline forgetting at 10 s could be explained only by the slowdown of recall, but one that is so strong as to require an even longer recall deadline beyond the 10 s used here. Indeed, as shown in Fig. 4, the cumulative recall rate of No-Think targets consistently increased as a function of time, and did not seem to reach a plateau. Thus, the forgotten targets at 10 s likely could have been recalled if participants had been given a much longer time to recall in the cued-recall test. If this is the case, the shapes of recall latency distributions would not be indicative of different mechanisms of the two strategies. Indeed, the situation of thought substitution where ones retrieve a part of items paired with the same cue corresponds to retrieval practice, which causes inhibition of items that share the cues with the retrieved items but not retrieved (retrieval induced forgetting; Anderson & Spellman, 1995; Anderson et al., 1994). That is, memory impairment by thought substitution is also explained by memory inhibition like direct suppression (e.g., Benoit & Anderson, 2012). In any case, the below-baseline forgetting at 10 s suggests that thought substitution leads to greater memory impairment than direct suppression.

The current results also illustrate that the trial duration of a recall test is crucial for detecting forgetting in a binary manner (i.e., recall rate). Given that many studies demonstrated below-baseline forgetting in recall tests with a duration of 4 s (Nardo & Anderson, 2024), 4 s can be accepted as a trial duration to find a sizable effect of retrieval stopping. Nevertheless, these characteristics of recall rates raise potential problems with the evaluation of individual differences in the ability of retrieval stopping. Many previous TNT studies defined the ability to control unwanted memories as the size of the below-baseline forgetting. Based on this definition, individuals who showed the same amount of below-baseline forgetting are thought to have an equivalent ability of memory control. However, there are still variances in the ability of memory control in terms of recall latency. In addition, the duration-dependent nature of the recall rate suggests that the size of below-baseline forgetting reflects the different memory control abilities among studies where the trial duration varies. This might make it difficult to compare their results concerning individual differences,

such as correlations between memory control ability and other variables. Taken together, recall latency can provide more fine-grained estimates of the effect of retrieval stopping and individual differences in memory control ability than recall rate.

It should be noted that participants' strategies in the cued-recall tests of the current study were possibly different from most of the previous TNT studies due to the deviation of the procedure that participants performed the task in a self-paced way by pressing the space bar. It might have encouraged participants to recall the targets as quickly as possible in the tests. Considering the speed–accuracy trade-off (Heitz, 2014), the speed-focused strategy might increase wrong first responses. However, in the current study, participants were allowed to correct their responses, and the corrected responses were counted as "successful recall." Therefore, the self-paced recall might have not critically influenced the recall rate results. In contrast, participants' strategies would influence recall latency per se in the final recall test. For example, the speed-focused strategy would lead to faster recall regardless of the conditions. However, their strategies presumably did not matter to the "slowdown" result of recall by retrieval stopping. Unless participants intentionally delayed target recall, it is not likely that any of their possible strategies (e.g., accuracy-focused strategy) would make recall latency so much longer that participants fail to recall targets within 10 s (i.e., "breakdown"). Taken together, the self-paced recall procedure used in the tests is unlikely to have had a critical impact on the current results.

Although there are several implications of the current study, there are two limitations in the results. First, the two experiments both failed to show below-baseline forgetting in the IP tests. Below-baseline forgetting in the IP test has been regarded as evidence of memory inhibition by retrieval stopping. Thus, it is still unclear that memory inhibition underlies the slowdown of memory recall. The present results in the IP tests might be due to different phonetic properties of letters between Japanese, which was used in the current study, and English, which was used in most of the previous studies showing below-baseline forgetting in the IP test. Each English (i.e., alphabetical) letter represents a single phoneme, whereas the majority of Japanese phonological letters represent larger single units, which consist of two phonemes (consonant + vowel). Thus, the first letters of targets presented along with independent probes have more information in Japanese than in English. This presumably made it easier for participants to recall corresponding words to a presented independent probe. Consistent with this reasoning, a previous study where the stimuli were English words failed to show below-baseline forgetting in an IP test when the first two letters of to-be-remembered items were presented as independent cues (Hulbert & Anderson, 2018, Experiment 1).

However, there also are several studies where below-baseline forgetting was not observed in the IP tests even though the materials were alphabetical words (e.g., Salamé & Danion, 2007; Wessel et al., 2005, 2020), suggesting that the phonemic issue is not the only factor that might make it difficult to detect below-baseline forgetting in the IP tests. Preexperimental semantic associations between the IP and target are obviously critical for performance on an IP test. It is likely that the associations were not controlled in the present materials. The problem is that there is no association norm that can be applied to materials for IP test cues across a variety of languages. Alternatively, no below-baseline forgetting in the current IP tests might indicate that little memory inhibition was induced by No-Think attempts in the current experiments.

Given the limitations of the IP test, other tasks can be utilized to test memory inhibition by retrieval stopping. When to-be-learned materials are words as in the current study, a word-naming task is a candidate. Healey et al. (2014) demonstrated slower naming of the words that were supposedly inhibited to resolve their interference in the preceded word generation task, compared with words in control conditions. When to-be-learned materials are images, a perceptual identification task is a candidate. Several TNT studies showed longer reaction times in perceptual identification of No-Think targets than Baseline ones (e.g., Gagnepain et al., 2014; Kim & Yi, 2013; Mary et al., 2020). Although these tasks do not evaluate forgetting directly, they are beneficial to examine memory inhibition as a potential factor of forgetting. In addition to averaging reaction times in the tasks, focusing on their distribution as in the current study might elucidate the nature of memory inhibition by retrieval stopping.

The second limitation was that the current "slowdown" results were possibly influenced by the criterion of the pre-TNT recall test. The current study required higher criterion (i.e., 80% accuracy) than most TNT studies (50% accuracy; cf., Nardo & Anderson, 2024). Although the strict criterion was introduced based on a theoretical consideration (see Method in detail), it presumably made cue–target associations stronger, and consequently, might reduce the likelihood of target memories being "broken down." Nevertheless, the issue raised in the current study remains important to consider in future studies with different training criteria: lower recall rates, which are often expressed as "forgetting," do not elucidate what changes in memory/recall processes underlie "forgetting." Suppression-induced "forgetting" is potentially misunderstood if it is interpreted in strong ways, such as loss or removal of memories (Davis et al., 2007).

In conclusion, the current experiments showed that retrieval stopping causes the slowdown of memory recall rather than the loss of item memory. Therefore, the binary classification of memory performance (recall success or

failure, i.e., recall rate) can lead to different results depending on trial duration in recall tests. These have been overlooked by excessive use of recall rate. Measuring recall latency requires researchers to estimate the onsets of vocal responses, which was a time- and effort-consuming task until recently, and this was one reason for the smaller number of chronometric studies on memory recall. However, onset estimation tools like Chronset (Roux et al., 2017) are now available, which help researchers measure recall latency. Measuring recall latency allows researchers to investigate cued recall process in more detail, which contributes to the theoretical development of memory research (cf. Osth & Farrell, 2019).

Funding This research was supported by a Grant-in-Aid for Science Research (B) from Japan Society for the Promotion of Science (20H01786) to S.S.; a Grant-in-Aid for JSPS Fellows (18J20948) to S.N.; and the Global Education Office, Kyoto University.

Declarations

Availability of data and materials All datasets for the current study are available at the Open Science Framework repository (https://osf.io/fuwhk/).

Code availability All analysis codes for the current study are available at the Open Science Framework repository (https://osf.io/fuwhk/).

Ethics approval The experiments reported in this article were approved by the Institutional Ethics Committee for Experimental Psychology Research at the Graduate School of Education, Kyoto University (approval number: CPE-203).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication Consent for publication was also obtained from all individual participants included in the study.

Conflicts of interest The authors have no competing interests to declare that are relevant to the content of this article.

Open practices statement The datasets for all experiments are available online (https://osf.io/fuwhk/). The materials for the experiments are available upon request. None of the experiments was preregistered.

References

Amano, S., & Kondo, T. (Eds.). (1999). *Nihongo-no goitokusei: Tango shinmitsudo* [Lexical properties of Japanese: Word familiarity] (Vol. 1). Sanseido.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063–1087. https://doi.org/10.1037/0278-7393.20.5.1063

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., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences, 18*(6), 279–292. https://doi.org/10.1016/j.tics.2014.03.002

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* (Vol. 58, pp. 53–120). Springer. https://doi.org/10.1007/978-1-4614-1195-6_3

Anderson, M. C., & Hulbert, J. C. (2021). Active forgetting: Adaptation of memory by prefrontal control. *Annual Review of Psychology*, *72*(1). https://doi.org/10.1146/annurev-psych-072720-094140

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

Anderson, M. C., Reinholz, J., Kuhl, B. A., & Mayr, U. (2011). Intentional suppression of unwanted memories grows more difficult as we age. *Psychology and Aging, 26*(2), 397–405. https://doi.org/10.1037/a0022505

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102*(1), 68–100. https://doi.org/10.1037/0033-295X.102.1.68

Audacity Team. (2015). *Audacity (R): Free audio editor and recorder* (Version 2.1.1) [Computer application]. Audacity. https://audacityteam.org/

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

Berkman, E. T., Burklund, L., & Lieberman, M. D. (2009). Inhibitory spillover: Intentional motor inhibition produces incidental limbic inhibition via right inferior frontal cortex. *NeuroImage, 47*(2), 705–712. https://doi.org/10.1016/j.neuroimage.2009.04.084

Blinch, J., & Cousineau, D. (2024). Response time distribution analysis of medium-sized datasets in MATLAB. *The Quantitative Methods for Psychology*, *20*(2), 106–120. https://doi.org/10.20982/tqmp.20.2.p106

Cano, M. E., & Knight, R. T. (2016). Behavioral and EEG evidence for auditory memory suppression. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00133

Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., & Anderson, M. C. (2015). Failing to forget: Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychological Science, 26*(5), 604–616.

Davis, M., Loftus, E. F., Rubin, D. C., & Wixted, J. T. (2007). Forgetting. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 315–337). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195310443.003.0015

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, 23*(4), 507–517. https://doi.org/10.1080/09658211.2014.907429

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via

targeted cortical inhibition. *Proceedings of the National Academy of Sciences, 111*(13), E1310–E1319. https://doi.org/10.1073/pnas.1311468111

Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel regulation of memory and emotion supports the suppression of intrusive memories. *Journal of Neuroscience, 37*(27), 6423–6441. https://doi.org/10.1523/JNEUROSCI.2732-16.2017

Haghighi, S. S., Ghorbani, M., Dehnavi, F., Safaie, M., & Moghimi, S. (2020). Motivated forgetting increases the recall time of learnt items: Behavioral and event related potential evidence. *Brain Research*, *1729*, Article 146624. https://doi.org/10.1016/j.brainres.2019.146624

Healey, M. K., Ngo, K. W. J., & Hasher, L. (2014). Below-baseline suppression of competitors during interference resolution by younger but not older adults. *Psychological Science, 25*(1), 145–151. https://doi.org/10.1177/0956797613501169

Heitz, R. P. (2014). The speed–accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00150

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review, 12*(3), 484–489.

Hulbert, J. C., & Anderson, M. C. (2018). What doesn't kill you makes you stronger: Psychological trauma and its relationship to enhanced memory control. *Journal of Experimental Psychology: General, 147*(12), 1931–1949. https://doi.org/10.1037/xge0000461

Kassambara, A. (2020). *rstatix: Pipe-friendly framework for basic statistical tests* [Computer software]. https://CRAN.R-project.org/package=rstatix

Kim, K., & Yi, D.-J. (2013). Out of mind, out of sight: Perceptual consequences of memory suppression. *Psychological Science, 24*(4), 569–574. https://doi.org/10.1177/0956797612457577

Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 35–45. https://doi.org/10.20982/tqmp.04.1.p035

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

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

Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., Vallée, T., . . . Gagnepain, P. (2020). Resilience after trauma: The role of memory suppression. *Science*, *367*(6479). https://doi.org/10.1126/science.aay8477

Nardo, D., & Anderson, M. C. (2024). Everything you ever wanted to know about the Think/No-Think task, but forgot to ask. *Behavior Research Methods, 56*, 3831–3860. https://doi.org/10.3758/s13428-024-02349-9

Nemeth, V. L., Kurgyis, E., Csifcsak, G., Maraz, A., Almasi, D. A., Drotos, G., . . . Must, A. (2014). The impact of intermediate-term alcohol abstinence on memory retrieval and suppression. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01396

Nishiyama, S., & Saito, S. (2021). Transferable inhibition of direct suppression: Evidence from a dot-probe task. *Psychologia, 63*(1), 20–41. https://doi.org/10.2117/psysoc.2020-A112

Nishiyama, S., & Saito, S. (2022). Retrieval stopping can reduce distress from aversive memories. *Cognition and Emotion, 36*(5), 957–974. https://doi.org/10.1080/02699931.2022.2071845

Osth, A. F., & Farrell, S. (2019). Using response time distributions and race models to characterize primacy and recency effects in free recall initiation. *Psychological Review, 126*(4), 578–609. https://doi.org/10.1037/rev0000149

Osth, A. F., Reed, A., & Farrell, S. (2021). How do recall requirements affect decision-making in free recall initiation? A linear ballistic accumulator approach. *Memory & Cognition, 49*, 968–983. https://doi.org/10.3758/s13421-020-01117-2

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org/

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods, 49*(5), 1864–1881. https://doi.org/10.3758/s13428-016-0830-1

Salamé, P., & Danion, J.-M. (2007). Inhibition of inappropriate responses is preserved in the think/no-think and impaired in the random number generation tasks in schizophrenia. *Journal of the International Neuropsychological Society*, *13*(02). https://doi.org/10.1017/S1355617707070300

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, 27*(5), 686–697. https://doi.org/10.1080/09658211.2018.1554079

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

Wang, Y., Cao, Z., Zhu, Z., Cai, H., & Wu, Y. (2015). Cue-independent forgetting by intentional suppression—Evidence for inhibition as the mechanism of intentional forgetting. *Cognition, 143*, 31–35. https://doi.org/10.1016/j.cognition.2015.05.025

Wang, Y., Luppi, A., Fawcett, J., & Anderson, M. C. (2019). Reconsidering unconscious persistence: Suppressing unwanted memories reduces their indirect expression in later thoughts. *Cognition, 187*, 78–94. https://doi.org/10.1016/j.cognition.2019.02.016

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, 28*(7), 870–887. https://doi.org/10.1080/09658211.2020.1797095

Wessel, I., Wetzels, S., Jelicic, M., & Merckelbach, H. (2005). Dissociation and memory suppression: A comparison of high and low dissociative individuals' performance on the Think-No Think task. *Personality and Individual Differences, 39*(8), 1461–1470. https://doi.org/10.1016/j.paid.2005.05.009

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer International Publishing. https://ggplot2.tidyverse.org

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.