Memory & Cognition
https://doi.org/10.3758/s13421-026-01910-5

Boosting hypermnesia with a cued-recall test format

Marilena L. Hirsch[1] · Oliver Kliegl[1] · Karl-Heinz T. Bäuml[1]

Received: 19 February 2026 / Accepted: 3 June 2026

Abstract

Recall performance of studied material typically increases across successive recall cycles, even in the absence of any corrective feedback. Prior research on this hypermnesia effect has predominantly employed free recall as the test format and has generally reported effects of moderate magnitude. Building on a recent incidental observation from a different experimental task, the present study examined whether an initial-letter cued recall (ILCR) format can amplify hypermnesia effects relative to a free-recall format. Across four experiments, participants repeatedly recalled previously studied item lists across three consecutive recall cycles. In Experiments 1-3, the test format involved either free recall (Experiments 2b and 3) or ILCR with a prescribed output order and a 6-s time limit per item (Experiments 1, 2a, and 3). Results showed larger hypermnesia effects in ILCR than in free recall, irrespective of whether words or images with their verbal labels served as study material and whether a 0-min delay or prolonged delays (30 min, 24 h) between study and test were applied. With both test formats, hypermnesia increased with delay. Experiment 4 showed that the results of Experiments 1-3 are critically tied to the constrained ILCR format and do not generalize to an ILCR condition in which all cues are presented simultaneously and participants are free to provide their responses in any order. Results align well with an interpretation based on both the changes in cue set account of hypermnesia and the retrieval practice hypothesis, suggesting a link between hypermnesia and the testing effect.

Keywords Hypermnesia · Multiple tests · Retrieval practice · Recall format · Delay

Hypermnesia refers to the finding that repeated attempts to retrieve previously studied information can lead to progressively improved recall across successive recall cycles. In a typical task examining this effect, participants are asked to study a list of items, usually words or pictures, before they complete several successive recall tests, each requiring them to produce the initially studied set of items. Item gains occur when items that were not produced earlier are recalled on a later test, whereas item losses are observed when items that were recalled initially are no longer retrieved on a later test. Hypermnesia arises when item gains outweigh item losses, leading to a net increase in the total number of items recalled over time. Conversely, when item losses exceed item gains, the result is net forgetting (for reviews, see Erdelyi, 1996; Payne, 1987).

Hypermnesia is a robust finding observed across a wide range of experimental settings. The effect has been found in numerous list learning paradigms, including those using unrelated words, associated word pairs, pictures, foreign language vocabulary, or nonsense syllables (e.g., Belmore, 1981; Kelley & Nairne, 2003; Mulligan, 2001; Roediger & Payne, 1982). Hypermnesia has also emerged in the recall of prose passages (Otani & Griffith, 1998; Wheeler & Roediger, 1985), in memory for films (Montangero, Ivanyi, & de Saint-Hilaire, 2003), in research on eyewitness memory (Dunning & Stern, 1992), and in studies of autobiographical memory (Bluck, Levine, & Laulhere, 1999).

While the hypermnesia finding is well established in the memory literature, the cognitive mechanisms underlying the effect remain debated. The three most prominent accounts are the changes in cue set account (Raaijmakers & Shiffrin, 1980; Roediger & Thorpe, 1978), the cumulative recall account (Roediger & Challis, 1989; Roediger, Payne, Gillespie, & Lean, 1982), and the retrieval strategy account (Erdelyi & Becker, 1974; Mulligan, 2001). According to the changes in cue set account, repeated recall attempts induce qualitative changes in the internal retrieval cues available at test. New cue sets are assumed to arise with newly recalled information, which may induce the use of different retrieval routes on a later test and thus enable recall of previously unrecalled information, thereby improving recall performance. In

✉ Marilena L. Hirsch
marilena.hirsch@ur.de

[1] Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany

Published online: 16 July 2026

contrast, the cumulative recall account assumes that initial recall attempts typically end before all potentially retrievable items have been recovered. Subsequent tests provide additional retrieval time, enabling access to items that were encoded but not yet recalled, thereby making conditions yielding higher overall recall levels more likely to exhibit hypermnesia. Finally, the retrieval strategy account attributes hypermnesia to qualitative improvements in retrieval organization across tests. Initial recall attempts are thought to foster the development of more efficient and structured retrieval plans, which facilitate re-access of previously recalled items and free cognitive resources for new retrievals, while simultaneously reducing item losses.

The magnitude of the hypermnesia effect is often modest, amounting to just two to five percentage points. Despite this fact, several variables have been found to influence the size of the effect. For instance, hypermnesia tends to be more pronounced when pictorial materials, such as line drawings, are used instead of verbal stimuli (Erdelyi & Becker, 1974). Similarly, longer time intervals between successive recall cycles (Mulligan, 2006) as well as longer delay intervals between the initial study phase and the later recall cycles (Wallner & Bäuml, 2018) seem to increase the magnitude of the effect. Nevertheless, the increase observed when moving from words to pictures, from short to longer intertest intervals, or from brief to extended delay intervals between study and the recall cycles again is typically moderate, falling in the range between two and five percentage points.

Against this background, a recent observation from a quite different experimental situation appears interesting, suggesting enhanced hypermnesia effects under certain conditions. Bäuml and Trißl (2022, Experiment 1) reported the results of a three-phase experiment on selective memory retrieval, in which participants in the first phase of the experiment studied a list of 15 unrelated words. In the second phase, which took place 30 min after the study and consisted of two successive practice cycles, participants in each cycle selectively retrieved ten predefined words of the study list. For this, the words' unique initial letters were provided as retrieval cues and were presented in a random order for 6 s each. In the third and final phase, memory for all 15 studied words was assessed, again using the words' unique initial letters as retrieval cues. When focusing on recall performance of the ten words that were practiced in the second phase of the experiment, the observation arose that recall performance of the words improved by approximately ten percentage points from the first to the second practice cycle, and improved by another approximately ten percentage points from the second practice cycle to the final recall test, indicating an unusually pronounced hypermnesia effect compared to the typically modest size of the effect found in typical hypermnesia studies.

Naturally, the experimental setup used in Bäuml and Trißl's (2022) selective retrieval experiment differed in a number of ways from the setup used in typical hypermnesia experiments. A first difference is the length of the study list. Whereas Bäuml and Trißl employed a study list consisting of 15 items, typical hypermnesia experiments have employed study lists of 40 items or even more (Mulligan, 2001; Roediger & Payne, 1982; Wallner & Bäuml, 2018). A second difference is that Bäuml and Trißl used a retrieval task in which, during selective retrieval, only a subset of the initially studied items were recalled, whereas in typical hypermnesia experiments, all initially studied items should be produced. As a third difference, Bäuml and Trißl administered the recall tests after a delay of 30 min, whereas in most hypermnesia experiments recall has occurred shortly after study. Because longer delays can increase the size of the hypermnesia effect (Wallner & Bäuml, 2018), part of the hypermnesia effect reported in Bäuml and Trißl may thus be due to longer delay. Finally, as a fourth difference, recall formats differ across studies. While recall format has mostly been free recall in earlier hypermnesia studies, in the Bäuml and Trißl study, an initial letter cued recall (ILCR) test constrained in output order and timing was used. In this test, the items' initial letters were used as retrieval cues and presented in a random order for 6 s each. Participants had to provide their response to each cue within the 6-s time frame.

Among the four listed differences between typical hypermnesia experiments and Bäuml and Trißl's (2022) selective retrieval experiment, the difference in recall format may be a particularly promising candidate to explain the observed difference in size of the hypermnesia effect between studies. Indeed, employing the cued-recall task used in Bäuml and Trißl's experiment may raise the possibility of larger item gains across repeated tests relative to the typical free-recall format. Increased item gains with an ILCR test constrained in output order and timing may emerge because, during a recall cycle, some participants may retrieve some studied items only after the items' initial letter cues have already disappeared from the screen and other list items should be recalled instead. In such case, participants will not overtly recall these retrieved items but rather may try to withhold the items until the items' cues reappear on the next recall cycle, thereby possibly inflating item gains on the subsequent recall cycle. In free recall, by contrast, such withholding is unnecessary, since participants can overtly report any newly retrieved item at any moment before the recall cycle ends.

ILCR tests constrained in output order and timing have been widely used in memory research, for instance, in research on the effects of selective memory retrieval (Anderson, Bjork, & Bjork, 1994; Bäuml and Trißl, 2022) or part-list cuing (Aslan & Bäuml, 2007; Bäuml & Aslan, 2004). Many real-world testing situations are also structured in ways that constrain performance during an initial recall attempt, for instance, by choosing a certain serial order for the retrieval of target memories or imposing time limits on single responses. Indeed, a student in an oral examination may retrieve relevant information only after the examiner has already proceeded to the next question, and the information not produced during the first attempt might well emerge if the examiner offered a second, later attempt. Surprisingly, there is not a single study in the literature to date that examined hypermnesia effects with such constrained tests, and, for instance, compared hypermnesia effects between an ILCR format constrained in output order and timing and a free-recall format. It is the goal of the present study to make a first step and, using conventional hypermnesia tasks, analyze hypermnesia effects both when using an ILCR format and a free-recall format at test.

The present study

This study pursued two primary objectives. The first objective was to replicate the finding reported in Bäuml and Trißl (2022, Experiment 1) within a typical hypermnesia setup. The issue was whether a relatively large hypermnesia effect – i.e., an effect in the order of 10–20 percentage points – can emerge after a prolonged delay between study and the recall tests when memory at test is assessed using Bäuml and Trißl's constrained ILCR format. A complementary aim was to determine whether any such increase in hypermnesia is specific to this constrained ILCR format or whether a similar pattern also arises with an unconstrained ILCR format in which all retrieval cues are presented simultaneously, and responses can be provided in any order. The second objective was to directly compare the magnitude of the hypermnesia effect when using the constrained ILCR format with the magnitude when using the more commonly employed free-recall format. Such comparisons were conducted both when using typical 0-min delays between study and the recall tests and when using prolonged delays, which enabled us to also examine whether delay increases the size of the hypermnesia effect not only with a free-recall (Wallner & Bäuml, 2018) but also in an ILCR format.

Four experiments were conducted. Experiment 1 examined whether repeated recall cycles produce a relatively large hypermnesia effect with words as study material, when there is a 30-min delay between study and the recall cycles and an ILCR format constrained in output order and timing at test. The goal was to replicate Bäuml and Trißl's (2022) finding in a more conventional hypermnesia setup. In addition, the experiment assessed whether the hypermnesia effect is larger after 30-min than a 0-min delay, allowing evaluation of whether the delay-related boost in hypermnesia that Wallner and Bäuml (2018) observed with a free-recall format generalizes to the ILCR format. Experiment 2a was aimed at replicating the results of Experiment 1 with images accompanied by their verbal labels as study material. Like in Experiment 1, the delay between study and the recall cycles was varied, and an ILCR format constrained in output order and timing was employed. Experiment 2b was identical to Experiment 2a but employed a free-recall format at test. A combined analysis of the results of Experiments 2a and 2b was planned to examine whether hypermnesia is indeed more pronounced with the cued-recall than the free-recall format. Experiment 3 provided a more direct comparison of hypermnesia effects between recall formats. Words were used as study material and the recall tests were conducted 24 h after study, either by using the constrained cued-recall format or a free-recall format. Two critical outcomes were anticipated for Experiments 1-3. First, on the basis of Bäuml and Trißl's (2022) observation, the results may show that the hypermnesia effect is more pronounced with the ILCR format than the free-recall format. Second, the hypermnesia effect may increase with delay not only with the free-recall but also with the ILCR format. In particular, the effects with cued recall may be larger than with free recall regardless of delay interval.

Finally, Experiment 4 examined whether the potential increase in hypermnesia observed with the ILCR format employed in Experiments 1–3 generalizes to an unconstrained ILCR format, in which all initial letter cues are presented simultaneously and participants may produce their responses in any order they prefer (see Fig. 1 for an overview of Experiments 1–4). To this end, the experiment compared the magnitude of hypermnesia between a free-recall format and this unconstrained ILCR format. The constraints on output order and timing implemented in Experiments 1–3 may have been critical for producing the expected ILCR-related increases in hypermnesia, as they could have contributed to the observed gains. When such constraints are removed, participants can provide responses at any moment during the test, thereby eliminating the possibility of responses being produced "too late" in the test. The results of Experiment 4 thus will indicate whether the constraints in output order and timing imposed on the ILCR tests of Experiments 1-3 contributed to the results of these experiments.

Experiment 1

Methods

Participants

Forty-four students of different German universities were recruited to participate in Experiment 1 (M = 23.8 years, SD = 2.8 years, 35 female, nine male, 0 diverse). The sample size was chosen to closely match that of Wallner and

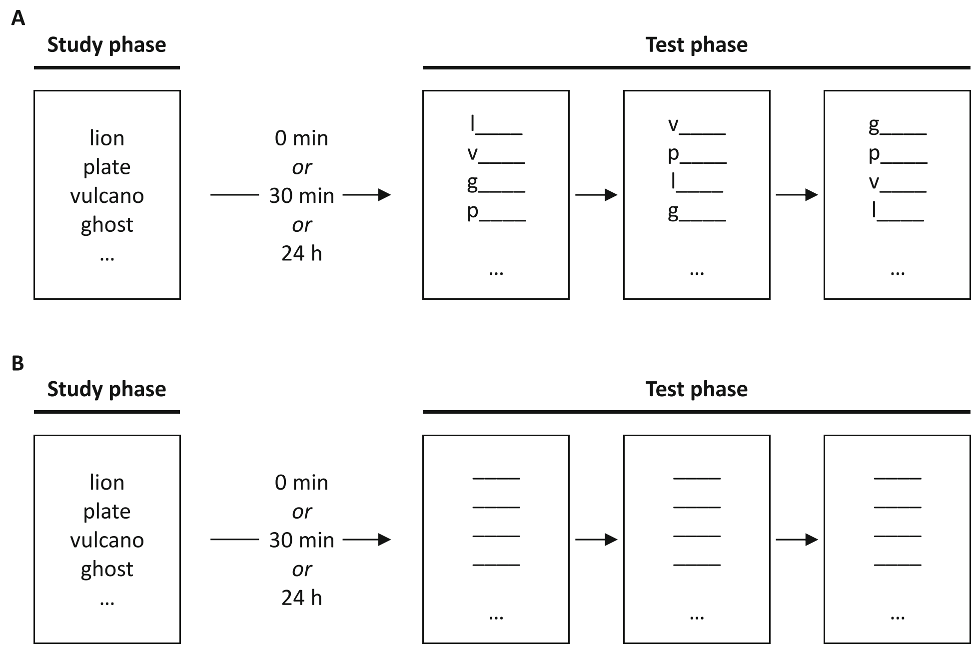

Fig. 1 Procedure of the experiments. *Note.* (**A**) Procedure in the initial letter cued-recall (ILCR) conditions (Experiments 1, 2a, 3, and 4): Participants studied a list of items. After delay intervals of 0 min, 30 min, or 24 h, they recalled the items in three successive recall cycles. The items' unique initial letters were provided as retrieval cues. While in Experiments 1, 2a, and 3, the single initial letter cues were provided successively, in Experiment 4, all initial letter cues were provided simultaneously and participants could produce their responses in any order they wished. (**B**) Procedure in the free-recall conditions (Experiments 2b, 3, and 4): Participants studied a list of items and after delay intervals of 0 min, 30 min, or 24 h, recalled the items in three successive recall cycles. No retrieval cues were provided, and participants were free to recall the items in any order they wished

Bäuml (2018, Experiment 1), an earlier study on hypermnesia which also employed an experimental design with two within-subject factors. All participants were native German speakers and were tested individually. Participation was compensated either with course credit or monetary payment.

Material Two lists containing 15 unrelated nouns each served as study materials. All nouns were high imagery items drawn from the norms of Snodgrass and Vanderwart (1980) and Severens, Van Lommel, Ratinckx, and Hartsuiker (2005). Within each list, every word began with a unique initial letter, which at test served as a retrieval cue for recall of the item. All items were translated into German.

Design The experiment employed a 2 × 3 repeated measures design with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 1, Test 2, Test 3). Participants recalled each word list after either a 0-min or 30-min delay across three successive cued-recall tests. Assignment of study lists and delay conditions was counterbalanced across participants.

Procedure The experiment was conducted via Zoom meetings (Zoom Video Communications, 2023), with the experimenter remaining present through audio and video for the entire session. After providing informed consent, participants completed two experimental blocks, each consisting of a study phase, a distractor task, and a test phase. The blocks differed in the length of delay (0 min vs. 30 min) between study and test. Block order was counterbalanced, with half of the participants starting with the 0-min delay condition and the other half starting with the 30-min delay condition. In each block, participants studied a list of 15 unrelated nouns presented individually in random order at a rate of 5 s per item. Immediately afterward, the list was presented again in a newly randomized order. Following the study phase, participants completed a 2-min backward counting task (subtracting three repeatedly from a three-digit number). In the 0-min delay condition, the test phase began directly after the counting task. In the 30-min delay condition, the counting task was followed by three 10-min distractor activities, consisting of puzzles targeting cognitive skills such as logical reasoning or mental rotation, creating a total delay of 30 min. The test phase comprised three successive cued-recall tests. In each test, participants attempted to recall as many studied items as possible. The initial letters of all target words were used as retrieval cues and were presented individually in random order for 6 s each. Participants were asked to provide their response within this 6-s interval. Participants responded orally, and the experimenter transcribed the answers. Between Test 1 and Test 2, as well as between Test 2 and Test 3, participants completed 3 min of arithmetic problems. After finishing the test phase of the first block, participants played Tetris for 5 min before beginning the second block (see Wallner & Bäuml, 2018, for a similar experimental setup).

Results

Net recall

Net recall performance for the 0-min and 30-min delay conditions across recall cycles is shown in Fig. 2. A 2 × 3 repeated measures ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 1, Test 2, Test 3) revealed a significant main effect of DELAY, $F(1, 43) = 9.76$, $MSE = 388.05$, $p = .003$, $\eta_p^2 = .19$, indicating that overall recall declined as the delay interval increased, consistent with typical time-dependent forgetting. The analysis also yielded a significant main effect of TEST, $F(1.60, 68.84) = 40.56$, $MSE = 90.26$, $p < .001$, $\eta_p^2 = .49$, demonstrating that net recall improved across successive recall cycles, a pattern reflecting hypermnesia. Critically, the DELAY × TEST interaction was significant, $F(2, 86) = 6.78$, $MSE = 53.72$, $p = .002$, $\eta_p^2 = .14$, showing that the size of the hypermnesia effect differed between delay conditions.

Examining the hypermnesia effect separately for each delay condition, two follow-up one-way ANOVAs with the factor of TEST were conducted. For the 0-min delay condition, the ANOVA revealed a significant effect of TEST, $F(1.65, 70.73) = 8.70$, $MSE = 86.07$, $p < .001$, $\eta_p^2 = .17$, confirming the presence of hypermnesia. Pairwise t-tests showed a significant increase in recall from Test 1 to Test 2, $t(43) = 2.91$, $p = .006$, Cohen's $d = 0.44$, but no significant increase from Test 2 to Test 3, $t(43) = 1.51$, $p = .137$, Cohen's $d = 0.23$. For the 30-min delay condition, ANOVA also revealed a significant effect of TEST, $F(1.72, 74.03) = 48.55$, $MSE = 64.10$, $p < .001$, $\eta_p^2 = .53$, revealing a larger hypermnesia effect than in the 0-min delay condition, which is consistent with the observed interaction. Follow-up comparisons showed significant increases in recall from Test 1 to Test 2, $t(43) = 6.06$, $p < .001$, Cohen's $d = 0.91$, as well as from Test 2 to Test 3, $t(43) = 4.07$, $p < .001$, Cohen's $d = 0.61$.

Table 1 Item gains and item losses in Experiment 1

Gains and losses	0 min *M*	0 min *SD*	30 min *M*	30 min *SD*
Gains Test 1 – Test 2	1.57	1.42	2.11	1.47
Gains Test 2 – Test 3	1.02	0.93	1.14	1.13
Losses Test 1 – Test 2	0.80	0.98	0.57	0.79
Losses Test 2 – Test 3	0.70	1.02	0.39	0.72

Item gains and losses

Table 1 shows item gains and item losses across recall cycles as a function of DELAY (0 min, 30 min). Item gains refer to

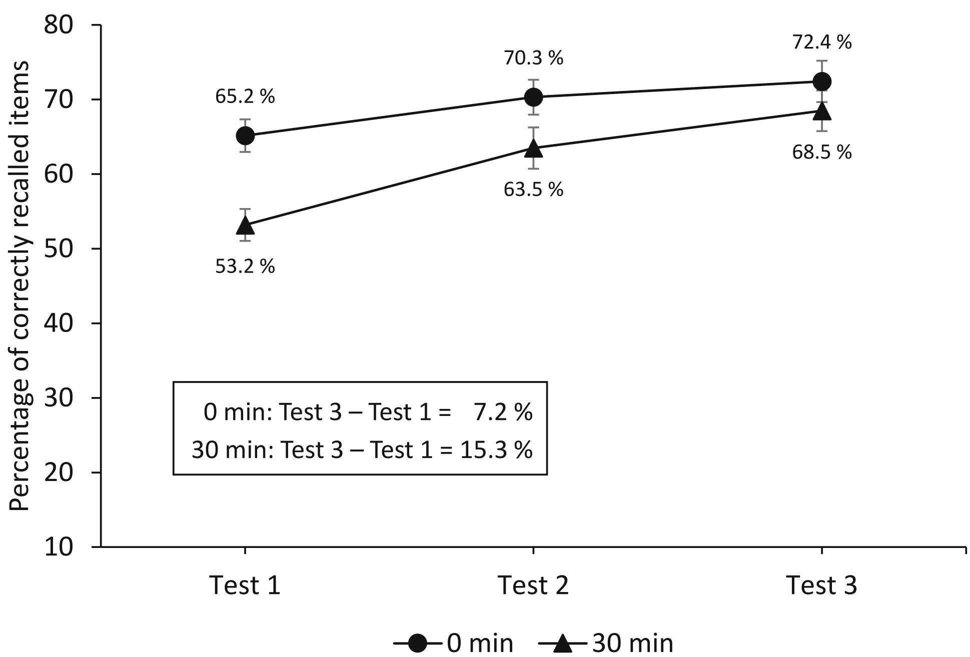

Fig. 2 Percentage of net recall in Experiment 1. *Note.* Percentage of net recall (ILCR test constrained in timing and output order) in the short delay and long delay conditions. *Error bars* represent standard errors

studied items recalled in Test 2 but not in Test 1, and studied items recalled in Test 3 but not in Test 2. Item losses refer to studied items recalled in Test 1 but not in Test 2, and studied items recalled in Test 2 but not in Test 3. For item gains, a 2 × 2 repeated measures ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 2, Test 3) revealed a significant main effect of TEST, $F(1, 43) = 14.39$, $MSE = 1.77$, $p < .001$, $\eta_p^2 = .25$, indicating that gains from Test 1 to Test 2 were larger than gains from Test 2 to Test 3. Neither the main effect of DELAY, $F(1, 43) = 3.48$, $MSE = 1.37$, $p = .069$, $\eta_p^2 = .08$, nor the DELAY × TEST interaction, $F(1, 43) = 1.71$, $MSE = 1.20$, $p = .198$, $\eta_p^2 = .04$, reached statistical significance. For item losses, the same ANOVA showed a significant main effect of DELAY, $F(1, 43) = 6.79$, $MSE = 0.48$, $p = .013$, $\eta_p^2 = .14$, indicating that more items were lost following the 0-min than the 30-min delay. Neither the main effect of TEST, $F(1, 43) = 1.52$, $MSE = 0.54$, $p = .225$, $\eta_p^2 = .03$, nor the interaction, $F(1, 43) < 1$, was statistically significant.

Discussion

Experiment 1 revealed a relatively large hypermnesia effect, reflected in an increase in net recall of approximately 15 percentage points across recall cycles. The finding held with words as study material and a constrained ILCR test after a prolonged delay of 30 min. It thus replicates the incidental observation made by Bäuml and Trißl (2022) in a selective retrieval task, who reported a hypermnesia effect of roughly 20 percentage points using the same recall format and the same 30-min delay condition (but a different word list for study). Going beyond Bäuml and Trißl's observation, Experiment 1 also demonstrated that the size of the hypermnesia effect increases with delay when this ILCR format is employed. Consistently, reliable hypermnesia effects were present after both delays, but the effect was larger after the 30-min delay than the 0-min delay, which generalizes Wallner and Bäuml's (2018) result that delay increases hypermnesia effects with the free-recall format. The increase in hypermnesia with delay was primarily driven by a reduction in item losses, which replicates results from Wallner and Bäuml's prior work.

The relatively large hypermnesia effects observed at both delays in Experiment 1 suggest that recall format may influence the overall size of the hypermnesia effect, given that prior studies using free recall typically reported more moderate effects, particularly at shorter delays. Experiments 2a and 2b addressed the possible role of recall format at test for hypermnesia. In these experiments, the same experimental setup was employed as in Experiment 1, but with two differences. First, both Experiment 2a and Experiment 2b used images and their verbal labels to assess whether the findings of Experiment 1 would generalize to this type of study material. The same stimuli were used in Experiments 2a and 2b. Second, whereas Experiment 2a followed Experiment 1 and employed the constrained ILCR format at test, in Experiment 2b a free-recall format was implemented at test. Because Experiments 2a and 2b were thus identical in setup with the only exception of recall format at test, we also conducted a combined analysis of the recall data of the two experiments to examine whether the cued-recall format leads to larger hypermnesia effects than the free-recall format and recall format moderates delay-related increases in hypermnesia.

Experiments 2a and 2b

Methods

Participants Following the rationale outlined in Experiment 1, 44 students of different German universities were recruited for Experiment 2a ($M = 23.9$ years, $SD = 3.6$ years, 38 female, six male, 0 diverse) and an additional 44 students of different German universities were recruited for Experiment 2b ($M = 24.1$ years, $SD = 3.3$ years, 33 female, 11 male, 0 diverse). All participants were native German speakers and were tested individually. Participation was compensated either with course credit or monetary payment.

Material, design, and procedure Like Experiment 1, both Experiments 2a and 2b were conducted via Zoom meetings. The experimental setup employed in Experiments 2a and 2b closely followed the setup used in Experiment 1, with three key modifications. First, images with their verbal labels rather than words only were used as study material. Two new lists of 15 items each were constructed, each consisting of simple black-and-white line drawings from the standardized picture norms of Snodgrass and Vanderwart (1980) and Severens et al. (2005). To ensure consistent item identification, which is critical for the cued-recall procedure employed in Experiment 2a, the verbal labels of each item were displayed beneath each drawing during study. Within each list, the verbal labels had unique initial letters. All items were translated into German. During the initial study phase, the images and their verbal labels were presented via a computer screen. Second, participants studied each list in only a single study cycle, in contrast to the two study cycles used in Experiment 1. This adjustment was implemented to reduce the risk of ceiling effects in the 0-min delay condition, because recall performance in Experiment 1 was rather high following the 0-min delay. Third, while Experiment 2a retained the cued-recall test format of Experiment 1, Experiment 2b employed a free-recall test format. With this format, participants were instructed in each recall cycle to orally recall as many of the studied items as possible in any order within 90 s, with the experimenter transcribing the answers. Total recall time in Experiment 2b (90 s) was matched to the cumulative recall time in Experiment 2a.

Results

Experiment 2a

Net recall

Net recall performance for the 0-min and 30-min delay conditions across recall cycles is shown in Fig. 3. A 2 × 3 repeated measures ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 1, Test 2, Test 3) revealed a significant main effect of DELAY, $F(1, 43) = 9.71$, $MSE = 421.86$, $p = .003$, $\eta_p^2 = .18$, confirming time-dependent forgetting. The analysis also yielded a significant main effect of TEST, $F(1.65, 70.73) = 36.91$, $MSE = 66.17$, $p < .001$, $\eta_p^2 = .46$, indicating that net recall increased across the three recall cycles. Most important, the DELAY × TEST interaction was significant, $F(1.70, 73.25) = 5.52$, $MSE = 40.68$, $p = .008$, $\eta_p^2 = .11$, demonstrating that the size of the hypermnesia effect varied with delay.

To further examine the pattern of hypermnesia, separate one-way repeated measures ANOVAs were conducted for each delay condition. For the 0-min delay condition, the ANOVA revealed a significant effect of TEST, $F(1.52, 65.51) = 13.44$, $MSE = 47.91$, $p < .001$, $\eta_p^2 = .24$, indicating robust hypermnesia. Pairwise comparisons showed significant increases in recall from Test 1 to Test 2, $t(43) = 3.71$, $p < .001$, Cohen's $d = 0.56$, and from Test 2 to Test 3, $t(43) = 2.49$, $p = .017$, Cohen's $d = 0.38$. For the 30-min delay condition, the corresponding ANOVA also revealed a significant main effect of TEST, $F(1.66, 71.33) = 32.51$, $MSE = 63.40$, $p < .001$, $\eta_p^2 = .43$, again demonstrating hypermnesia, but with a larger effect size than after the 0-min delay. Pairwise analyses confirmed significant increases in recall from Test 1 to Test 2, $t(43) = 5.11$, $p < .001$, Cohen's $d = 0.77$, and from Test 2 to Test 3, $t(43) = 3.57$, $p < .001$, Cohen's $d = 0.54$.

Table 2 Item gains and item losses in Experiment 2a and Experiment 2b

Gains and losses	0 min *M*	0 min *SD*	30 min *M*	30 min *SD*
Experiment 2a				
Gains Test 1 – Test 2	1.41	1.06	1.77	1.33
Gains Test 2 – Test 3	1.07	1.04	0.93	1.02
Losses Test 1 – Test 2	0.86	1.00	0.57	0.76
Losses Test 2 – Test 3	0.61	0.78	0.30	0.59
Experiment 2b				
Gains Test 1 – Test 2	0.34	0.57	0.57	0.66
Gains Test 2 – Test 3	0.45	0.63	0.34	0.68
Losses Test 1 – Test 2	0.55	0.73	0.18	0.45
Losses Test 2 – Test 3	0.30	0.51	0.20	0.51

Item gains and losses

Table 2 shows item gains and item losses for Experiment 2a across recall cycles as a function of DELAY (0 min, 30 min).

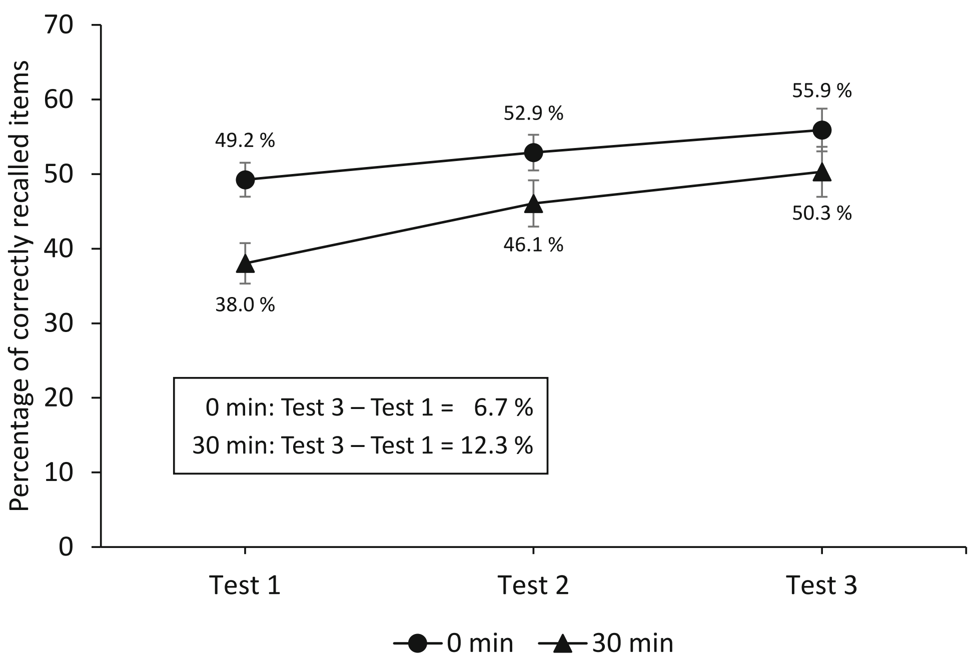

Fig. 3 Percentage of net recall in Experiment 2a. *Note.* Percentage of net recall (ILCR test constrained in timing and output order) in the short delay and long delay conditions. *Error bars* represent standard errors

For item gains, a 2 × 2 ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 2, Test 3) showed a main effect of TEST, $F(1, 43) = 13.31$, $MSE = 1.15$, $p < .001$, $\eta_p^2 = .24$, indicating more item gains from Test 1 to Test 2 than from Test 2 to Test 3, but no main effect of DELAY, $F(1, 43) < 1$, and no interaction between factors, $F(1, 43) = 3.17$, $MSE = 0.87$, $p = .082$, $\eta_p^2 = .07$. Regarding item losses, the same ANOVA revealed significant main effects of DELAY, $F(1, 43) = 6.34$, $MSE = 0.65$, $p = .016$, $\eta_p^2 = .13$, and TEST, $F(1, 43) = 6.08$, $MSE = 0.49$, $p = .018$, $\eta_p^2 = .12$, reflecting more losses for the 0-min than the 30-min delay and more losses from Test 1 to Test 2 than from Test 2 to Test 3. The interaction between factors was not significant, $F(1, 43) < 1$.

Experiment 2b

Net recall

Net recall performance for the 0-min and 30-min delay conditions across test cycles is shown in Fig. 4. A 2 × 3 ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 1, Test 2, Test 3) showed a significant main effect of DELAY, $F(1, 43) = 16.42$, $MSE = 590.68$, $p < .001$, $\eta_p^2 = .28$, reflecting time-dependent forgetting, but no significant main effect of TEST, $F(1.74, 74.63) = 3.04$, $MSE = 21.51$, $p = .061$, $\eta_p^2 = .07$. Critically, there was a reliable interaction between the two factors, $F(2, 86) = 7.47$, $MSE = 14.68$, $p = .001$, $\eta_p^2 = .15$, indicating that delay modulated the size of the hypermnesia effect. For the 0-min delay, a follow-up one-way ANOVA with the within-subject factor of TEST (Test 1, Test 2, Test 3) revealed no effect of TEST, $F(1.75, 75.11) = 1.10$, $MSE = 23.46$, $p = .331$, $\eta_p^2 = .03$, indicating no hypermnesia. In contrast, for the 30-min delay, the same ANOVA revealed an effect of TEST, $F(2, 86) = 11.19$, $MSE = 12.85$, $p < .001$, $\eta_p^2 = .21$, reflecting reliable hypermnesia. Follow-up pairwise comparisons showed an increase in recall from Test 1 to Test 2, $t(43) = 3.72$, $p < .001$, Cohen's $d = 0.56$, but no increase in recall from Test 2 to Test 3, $t(43) = 1.23$, $p = .225$, Cohen's $d = 0.19$.

Item gains and losses

Table 2 shows item gains and item losses for Experiment 2b across recall cycles as a function of DELAY (0 min, 30 min). Regarding item gains, a 2 × 2 ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 2, Test 3) showed no main effects of DELAY, $F(1, 43) < 1$, and TEST, $F(1, 43) < 1$, and no significant interaction between factors, $F(1, 43) = 3.34$, $MSE = 0.38$, $p = .075$, $\eta_p^2 = .07$. For item losses, the same ANOVA showed a significant main effect of DELAY, $F(1, 43) = 6.64$, $MSE = 0.34$, $p = .014$, $\eta_p^2 = .13$, reflecting more losses after the 0-min than the 30-min delay. Neither the main effect of TEST, $F(1, 43) = 1.97$, $MSE = 0.29$, $p = .168$, $\eta_p^2 = .04$, nor the interaction between factors, $F(1, 43) = 3.15$, $MSE = 0.26$, $p = .083$, $\eta_p^2 = .07$, was statistically significant.

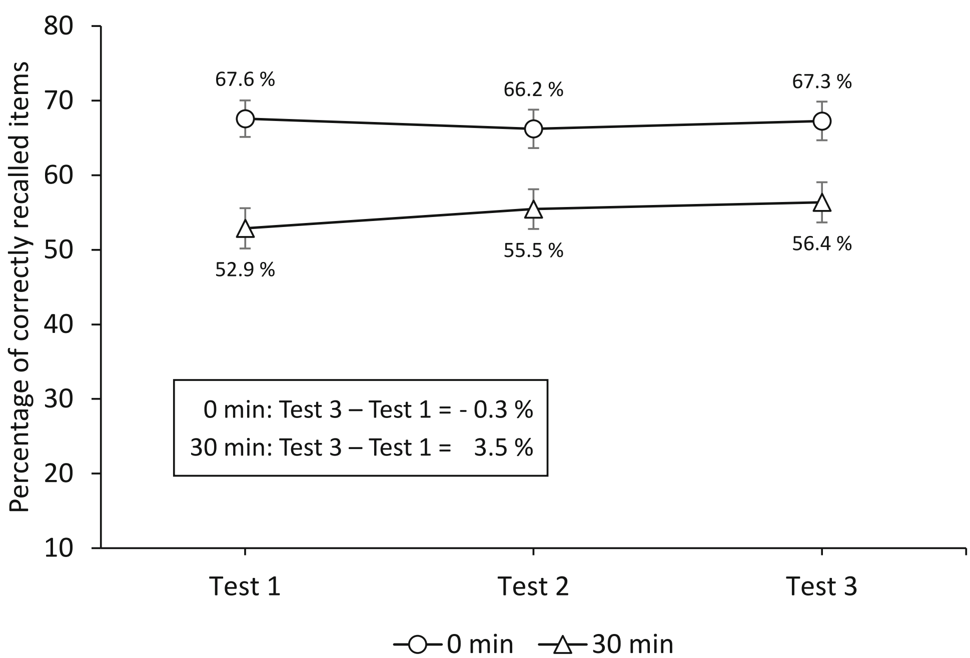

Fig. 4 Percentage of net recall in Experiment 2b. *Note.* Percentage of net recall (free recall) in the short-delay and long-delay conditions. *Error bars* represent standard errors

Combined analysis of Experiments 2a and 2b

Net recall

Including the recall data of both experiments, a 2 × 3 × 2 ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 1, Test 2, Test 3) and the between-subjects factor of TEST FORMAT (ILCR, free recall) showed significant main effects of DELAY, TEST, and TEST FORMAT, all Fs ≥ 15.01, all ps < .001, all η_p^2s ≥ .15, reflecting that overall recall performance was greater after the 0-min than the 30-min delay, increased across recall cycles, and was greater with the free-recall than the ILCR format. There were also significant two-way interactions between DELAY and TEST and between TEST and TEST FORMAT, all Fs ≥ 11.81, all ps < .001, all η_p^2s ≥ .12, reflecting that overall recall performance increased less across test cycles after the 0-min than the 30-min delay, and, most important, that the hypermnesia effect was overall more pronounced with the ILCR than the free-recall format. Neither the two-way interaction of DELAY and TEST FORMAT, nor the three-way interaction reached statistical significance, all Fs ≤ 1.17, all ps ≥ .282, all η_p^2s ≤ .01.

Item gains and losses

For item gains, a 2 × 2 × 2 ANOVA with the within-subject factors of DELAY (0 min, 30 min) and TEST (Test 2, Test 3) and the between-subjects factor of TEST FORMAT (ILCR, free recall) showed no main effect of DELAY, $F(1, 86) < 1$, but significant main effects of TEST and TEST FORMAT, all Fs ≥ 12.36, all ps < .001, all η_p^2s ≥ .13, reflecting that overall item gains were greater between Test 1 and Test 2 than between Test 2 and Test 3, and were greater in ILCR than free recall. There were also significant two-way interactions between DELAY and TEST and between TEST and TEST FORMAT, all Fs ≥ 6.23, all ps ≤ .014, all η_p^2s ≥ .07, reflecting that the overall reduction in item gains across test cycles was greater after the 30-min than the 0-min delay and that there was a more pronounced decrease overall in item gains across test cycles with the ILCR than the free-recall format. There was no significant interaction between DELAY and TEST FORMAT and no significant three-way interaction, all Fs < 1, all ps ≥ .638, all η_p^2s < .01. For item losses, the same three-way ANOVA showed significant main effects of DELAY, TEST, and TEST FORMAT, all Fs ≥ 7.90, all ps ≤ .006, all η_p^2s ≥ .08, reflecting that overall item losses were greater after the 0-min than the 30-min delay, decreased across recall cycles, and were greater in ILCR than free recall. None of the interactions reached statistical significance, all Fs ≤ 1.62, all ps ≥ .206, all η_p^2s ≤ .02.

Discussion

Using images accompanied by verbal labels instead of words only as study material, the results of Experiment 2a replicate the results observed in Experiment 1 by demonstrating hypermnesia after both the 0-min and the 30-min delays when using the constrained ILCR format. Importantly, like in Experiment 1, the size of the hypermnesia effect increased with delay. Using the free-recall format, Experiment 2b also showed a reliable hypermnesia effect after the 30-min delay, but, in contrast to Experiment 2a, no reliable hypermnesia effect after the 0-min delay. In both Experiment 2a and Experiment 2b, the delay-related increase in hypermnesia was driven by reduced item losses, which mirrors the pattern observed in Experiment 1 and aligns with the free-recall findings of Wallner and Bäuml (2018). The present finding of a stronger hypermnesia effect after prolonged delay that is mainly driven by a reduction in item losses also fits with the assumption that retrieval practice reduces the forgetting of practiced items (Hogan & Kintsch, 1971; Roediger & Karpicke, 2006) and does so to an even greater degree if retrieval is demanding (Bäuml, Holterman, & Abel, 2014; Carpenter, 2011), i.e., after the prolonged delay.

Experiments 2a and 2b also enabled a more direct evaluation of the effects of recall format on hypermnesia. A combined analysis of the recall data from the two experiments revealed that – regardless of delay – cued recall produced a larger hypermnesia effect than free recall. This difference was driven by greater overall item gains in cued recall compared to free recall. Somewhat unexpectedly, overall item losses were also higher in ILCR than in free recall, but the difference between recall formats was smaller than the one observed for item gains, resulting in a larger hypermnesia effect with the ILCR format.

The absence of a significant hypermnesia effect after the 0-min delay with free recall in Experiment 2b contrasts with other findings in the literature, including the findings of Wallner and Bäuml (2018), who observed a significant hypermnesia effect following a similarly brief delay of 3 min when using pictorial materials. The difference in findings might have to do with differences in study material. The present study used other study items than the prior studies and, in particular, employed a much shorter list of 15 items only, which might influence the size of the hypermnesia effect (but see General discussion). On the other hand, hypermnesia effects after short delay are often small in size (see above), and finding no significant effect of recall cycles in a single experiment certainly is within the range of possible findings.

Experiments 2a and 2b provide insights into the possible role of recall format in shaping hypermnesia, but caution is warranted because insights rely on a between-experiments comparison. It was thus the aim of Experiment 3 to more directly assess the potential influence of recall format on hypermnesia. For this, a within-experiment manipulation of recall format was conducted, allowing more clearcut conclusions on whether the ILCR format can lead to larger hypermnesia effects than the free-recall format.

Experiment 3

Methods

Participants Because Experiment 3 employed a mixed design with one within-subject and one between-subjects factor, we conducted an a priori power analysis to determine the required sample size using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). The analysis indicated that a sample of 33 participants per between subjects condition would be required to detect a small to medium-sized interaction effect of f = 0.20 with α = .05 and 1 - β = .95. Based on this recommendation and on the sample sizes used in prior hypermnesia studies (Mulligan, 2006; Wallner & Bäuml, 2018), 88 participants of different German universities were sampled overall (*M* = 24.8 years, *SD* = 3.2 years, 67 female, 21 male, 0 diverse), with 44 participants in the cued-recall condition and 44 participants in the free-recall condition. All participants were native German speakers and were tested individually. Participation was compensated either with course credit or monetary payment.

Material, design, and procedure Like the preceding experiments, Experiment 3 was conducted via Zoom Meetings. The study materials used in Experiment 3 were identical to those of Experiment 1, with assignment of word lists to the two test format conditions counterbalanced across participants. The experiment employed a 3 × 2 mixed design with the within-subject factor of TEST (Test 1, Test 2, Test 3) and the between-subjects factor of TEST FORMAT (ILCR, free recall). The two test formats were implemented in the same way as in Experiments 2a and 2b, but with three modifications. First, all participants completed only one experimental block with a prolonged 24-h delay between study and test. For each participant, the study phase occurred in Session 1, and the test phase in Session 2, 1 day later. Second, similar to Experiments 2a and 2b, the word list was presented once only, but the presentation duration was increased to 6 s per item to ensure sufficiently high recall levels after the 24-h delay. Third, following the 2-min backward counting distractor task, participants completed another 4-min arithmetic task to further reduce the chances that they would engage in recall of the word list immediately after the end of Session 1.

Results

Net recall

Net recall performance for the cued-recall and free-recall conditions across recall cycles is shown in Fig. 5. A 3 × 2

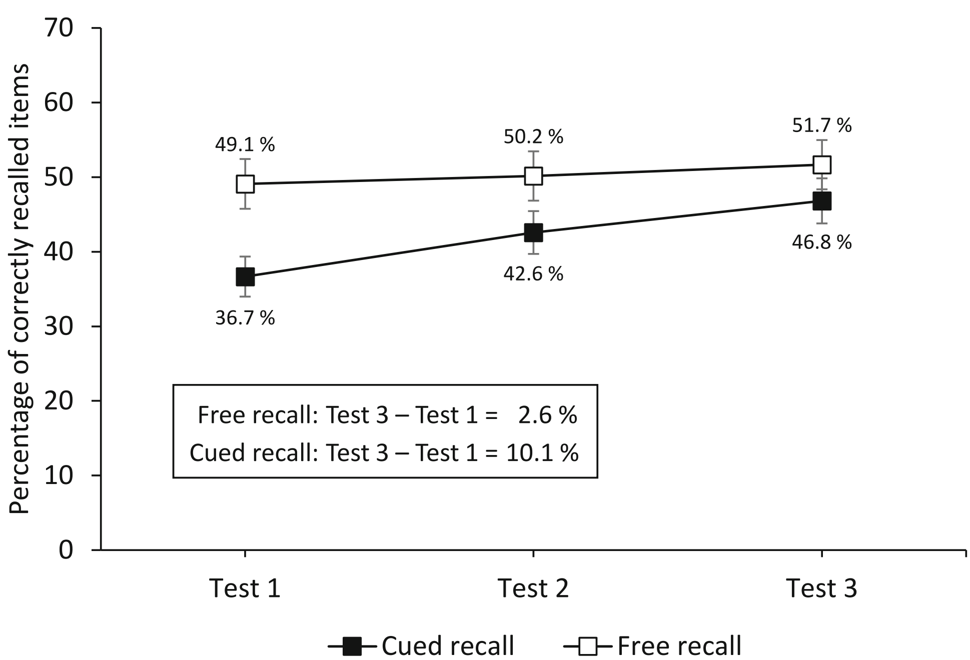

Fig. 5 Percentage of net recall in Experiment 3. *Note.* Percentage of net recall in the ILCR test constrained in timing and output-order and free-recall test after the 24-h delay. *Error bars* represent standard errors

ANOVA with the within-subject factor of TEST (Test 1, Test 2, Test 3) and the between-subjects factor of TEST FORMAT (ILCR, free recall) showed a main effect of TEST $F(1.51, 129.65) = 39.35$, $MSE = 30.12$, $p < .001$, $\eta_p^2 = .31$, reflecting hypermnesia, but no main effect of TEST FORMAT, $F(1, 86) = 3.71$, $MSE = 1220.64$, $p = .057$, $\eta_p^2 = .04$. Critically, the interaction between TEST and TEST FORMAT was significant, $F(1.51, 129.65) = 14.26$, $MSE = 30.12$, $p < .001$, $\eta_p^2 = .14$, showing that the size of the hypermnesia effect varied with test format. For ILCR, a one-way ANOVA with the within-subject factor of TEST revealed a significant main effect of TEST, $F(1.39, 59.54) = 33.27$, $MSE = 49.66$, $p < .001$, $\eta_p^2 = .44$, showing hypermnesia. Follow-up t tests revealed a significant increase in recall from both Test 1 to Test 2, $t(43) = 5.13$, $p < .001$, Cohen's $d = 0.77$, and Test 2 to Test 3, $t(43) = 4.73$, $p < .001$, Cohen's $d = 0.71$. For free recall, the same ANOVA also showed a significant main effect of TEST, $F(1.73, 74.46) = 6.68$, $MSE = 12.74$, $p = .003$, $\eta_p^2 = .14$, but with a substantially smaller effect size than for cued recall. Follow-up t tests showed no difference in recall between Tests 1 and 2, $t(43) = 1.86$, $p = .070$, Cohen's $d = 0.28$, but significantly higher recall in Test 3 than Test 2, $t(43) = 2.12$, $p = .040$, Cohen's $d = 0.32$.

Item gains and losses

Table 3 shows item gains and item losses across recall cycles as a function of TEST FORMAT (ILCR, free recall). Regarding item gains, a 2 × 2 ANOVA with the within-subject factor of TEST (Test 2, Test 3) and the between-subjects factor of TEST FORMAT showed a significant main effect of TEST FORMAT, $F(1, 86) = 30.84$, $MSE = 0.98$, $p < .001$, $\eta_p^2 = 0.26$, with overall higher gains in ILCR than free recall, but no main effect of TEST, $F(1, 86) = 3.26$, $MSE = 0.50$, $p = .075$, $\eta_p^2 = 0.04$. Critically, the interaction between TEST and TEST FORMAT was significant, $F(1, 86) = 4.97$, $MSE = 0.50$, $p = .028$, $\eta_p^2 = .06$, suggesting that test format modulated item gains. Follow-up pairwise comparisons showed significantly higher item gains in Test 2 than Test 3 in cued recall, $t(43) = 2.45$, $p = .018$, Cohen's $d = 0.37$, but not in free recall, $t(43) < 1$. Regarding item losses, the same ANOVA showed a significant main effect of TEST FORMAT, $F(1, 86) = 9.96$, $MSE = 0.30$, $p = .002$, $\eta_p^2 = .10$, reflecting more losses in cued recall than free recall. There was no main effect of TEST, $F(1, 86) = 2.51$, $MSE = 0.18$, $p = .117$, $\eta_p^2 = .03$, and no interaction between factors, $F(1, 86) = 1.52$, $MSE = 0.18$, $p = .221$, $\eta_p^2 = .02$.

Table 3 Item gains and item losses in Experiment 3

	Cued recall		Free recall	
Gains and losses	*M*	*SD*	*M*	*SD*
Gains Test 1 – Test 2	1.34	1.18	0.27	0.50
Gains Test 2 – Test 3	0.91	0.96	0.32	0.64
Losses Test 1 – Test 2	0.45	0.66	0.11	0.32
Losses Test 2 – Test 3	0.27	0.59	0.09	0.29

Discussion

Experiment 3 revealed a robust hypermnesia effect in both the ILCR and the free-recall condition after a 24-h delay. Critically, the direct comparison of recall formats showed that the size of the hypermnesia effect was larger in cued than free recall, thereby corroborating the results from the combined analysis of Experiments 2a and 2b. When considering item gains and item losses, the larger hypermnesia effect in the ILCR than in the free-recall condition was again attributable to larger item gains. Item losses were also again higher in the ILCR than in the free-recall condition, but the increase in item gains outweighed the increase in item losses, resulting in a net advantage for the ILCR format, which replicates the finding of Experiments 2a and 2b.

The aim of Experiment 4 was to assess whether the results of the prior experiments obtained with ILRC tests constrained in timing and output order generalize to an unconstrained ICLR format. Therefore, the same experimental procedure was used as in Experiment 3, except that the initial letters in the ILCR condition were all presented simultaneously rather than successively. Participants were thus able to recall items in any order and without time constraints for individual items. If constraints in timing and output order were critical for the pronounced hypermnesia effects in the ILCR conditions of the present Experiments 1–3, then presenting initial letter cues simultaneously should lead to a hypermnesia effect whose size more closely aligns with that in free recall.

Experiment 4

Methods

Participants Following Experiment 3, a total of 88 participants from different German universities were recruited to take part in Experiment 4 (M = 22.6 years, SD = 2.5, 69 female, 18 male, one diverse), with 44 participants in the cued-recall condition and 44 participants in the free-recall condition. All participants were native German speakers and were tested individually. Participation was compensated either with course credit or monetary payment.

Materials, design, and procedure Like in the preceding experiments, Experiment 4 was conducted via Zoom meetings. All procedural details were identical to those of Experiment 3, with two exceptions. First, in the ILCR condition, initial letter cues were presented simultaneously in a new random order on each test. Participants were instructed that they had 90 s to orally recall as many items as possible in any order and were encouraged to use the initial letter cues to guide memory. Thus, the ILCR test in Experiment 4 was effectively unconstrained, as participants were free to determine the order in which items were recalled and could, in principle, devote more than the 6 s allotted in the earlier experiments using the constrained ILCR format to retrieve individual items.

Second, the line drawings together with their corresponding verbal labels that had been used as study materials in Experiment 3 were presented to only a subset of participants, i.e., 24 participants in each of the ILCR and free-recall conditions. The remaining 20 participants in each test format instead studied the word stimuli that had been used in Experiment 1.

Results

Net recall

Net recall performance for the cued-recall and free-recall conditions across tests is shown in Fig. 6. A 3 × 2 ANOVA with the within-subject factor TEST (Test 1, Test 2, Test 3) and the between-subjects factor TEST FORMAT (ILCR vs. free recall) revealed no significant main effect of TEST FORMAT, $F(1, 86) = 2.15$, $MSE = 1202.35$, $p = .146$, $\eta_p^2 = .02$, but a significant main effect of TEST, $F(1.78, 153.02) = 9.71$, $MSE = 15.92$, $p < .001$, $\eta_p^2 = .10$, indicating an overall increase in recall performance across tests, reflecting hypermnesia. Follow-up t-tests showed a significant increase in recall from Test 1 to Test 2, $t(87) = 2.31$, $p = .023$, $d = 0.25$, as well as from Test 2 to Test 3, $t(87) = 2.54$, $p = .013$, $d = 0.27$. Critically, the interaction between TEST and TEST FORMAT did not reach significance, $F(1.78, 153.02) < 1$, suggesting that hypermnesia was similar in magnitude for free-recall and unconstrained ILCR.

Table 4 Item gains and item losses in Experiment 4

Gains and losses	Cued recall *M*	Cued recall *SD*	Free recall *M*	Free recall *SD*
Gains Test 1 – Test 2	0.48	0.76	0.32	0.67
Gains Test 2 – Test 3	0.41	0.73	0.18	0.39
Losses Test 1 – Test 2	0.27	0.50	0.14	0.35
Losses Test 2 – Test 3	0.16	0.43	0.07	0.25

Item gains and losses

Table 4 shows item gains and item losses in the cued-recall and free-recall conditions. Two 2 × 2 ANOVAs with the within-subject factor TEST (Test 2, Test 3) and the between-subjects factor TEST FORMAT (ILCR vs. free recall) revealed no significant main effects and interactions for either item gains, all $Fs \leq 3.16$, all $ps \geq .079$, all $\eta_p^2 s \leq .04$, or item losses, all $Fs \leq 3.50$, all $ps \geq .065$, all $\eta_p^2 s \leq .04$.

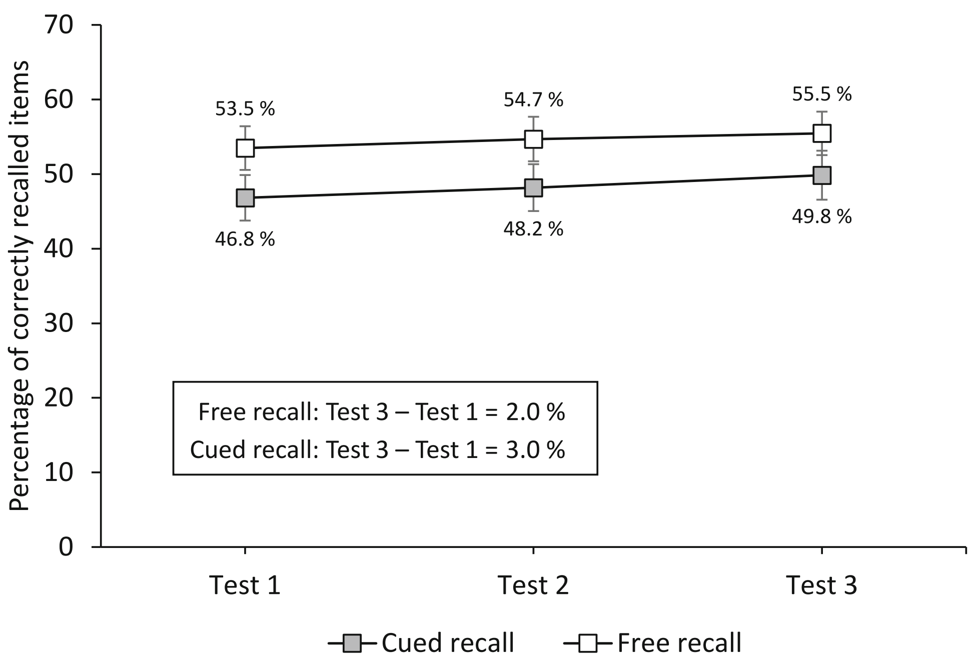

Fig. 6 Percentage of net recall in Experiment 4. *Note.* Percentage of net recall in the unconstrained ILCR test and free-recall test after the 24-h delay. *Error bars* represent standard errors

Discussion

The results of Experiment 4 demonstrate hypermnesia effects of comparable magnitude for unconstrained ILCR and free recall. These findings indicate that the constraints on timing and output order imposed on ILCR in Experiments 1–3 were critical for producing the larger hypermnesia effects demonstrated in those earlier experiments relative to free recall. Consistent with this interpretation, the convergence in recall performance between unconstrained ILCR and free recall observed in Experiment 4 also extended to item-level dynamics. Specifically, unlike in Experiment 3, item gains and item losses in the ILCR condition of Experiment 4 did not exceed those observed in the free-recall condition. A final critical difference relative to Experiment 3 concerns Test 1 recall levels. In Experiment 4, recall performance in the unconstrained ILCR format was only slightly lower numerically on Test 1 than in free recall (6.7 percentage points), whereas in Experiment 3, this difference was quite pronounced (12.4 percentage points). The pattern found in Experiment 4 closely resembles results reported by Carpenter and DeLosh (2006), who likewise observed only small reductions in recall performance for unconstrained ILCR relative to free recall.

General discussion

The findings of the present study demonstrate that usage of an ILCR format constrained in output order and timing during recall cycles can elicit a pronounced hypermnesia effect. Using words as study material and a prolonged 30-min delay interval between study and test, recall performance in Experiment 1 increased by 15.3 percentage points from the first to the third recall cycle. This effect exceeds the size of the hypermnesia effect typically reported in the literature and generalizes Bäuml and Trißl's (2022) observation from a selective retrieval task to a more conventional hypermnesia task. The results of Experiments 2a and 3 extend the finding by showing similarly sized hypermnesia effects with initial letter cues after a 30-min delay when picture stimuli accompanied the verbal material (12.3 percentage points, Experiment 2a), and when again words only were used as study material but delay was increased to 24 h (10.1 percentage points, Experiment 3). In all these cases, the hypermnesia effects were primarily driven by substantial item gains across recall cycles.

The results of this study also show that the size of the hypermnesia effect can be larger when an ILCR test constrained in output order and timing is employed at test compared to a free-recall format. This finding arose both in a post-hoc comparison between Experiment 2a (cued recall) and Experiment 2b (free recall) after delays of both 0 min and 30 min, and in a direct within-experiment comparison of the two recall formats in Experiment 3 after a delay of 24 h. The analyses further revealed that hypermnesia increased with delay both when using free recall and when using the constrained ILCR format, which was primarily driven by reduced item losses across recall cycles. While the finding that hypermnesia increases with delay with free-recall (Experiment 2b) replicates earlier results reported by Wallner and Bäuml (2018), the delay-induced increase in hypermnesia with ILCR (Experiment 1, Experiment 2a) is a new finding. Overall, the findings show that both recall format and delay can critically influence the size of the hypermnesia effect.

Importantly, the results of Experiment 4 additionally demonstrate that usage of an unconstrained ILCR test can lead to a hypermnesia effect whose size closely matches that in free recall (3.0 vs. 2.0 percentage points), suggesting that the finding of increased hypermnesia with the constrained ILCR format does not generalize to unconstrained ILCR. This outcome aligns with findings from an earlier study by Payne, Hembrooke, and Anastasi (1993), which likewise reported hypermnesia effects of comparable magnitude for free-recall and an unconstrained cued-recall format. Payne et al. employed categorized word lists for study and provided category names as cues during test cycles but, like in the present Experiment 4, allowed participants to recall the studied items in any order they wished. The results by Payne et al. thus converge with the present results, supporting the view of a critical role of retrieval constraints for the size of the hypermnesia effect.

The present finding that hypermnesia is affected by the degree to which retrieval restrictions are present during tests aligns well with an interpretation of the changes in cue set account of hypermnesia (Raaijmakers & Shiffrin, 1980; Roediger & Thorpe, 1978). The changes in cue set account can explain why, relative to free recall, the size of the hypermnesia effect can be larger when an ILCR test constrained in output order and timing is used during recall cycles. With this test format, participants – within a recall cycle – may miss recall of some early tested items just to retrieve the items after the items' first letter cues have already disappeared from the screen. Such too-late retrieval may modify participants' internal cue set used on the next recall cycle and increase the probability that these items will be retrieved when their initial letter cues reappear on the next recall cycle. This leads to the prediction that item gains across successive recall cycles should be increased, relative to a free-recall format, which is what the results of Experiments 1-3 show. Indeed, when no restrictions are imposed on ILCR tests, the size of the hypermnesia effect closely matches that observed in free recall, as is shown in Experiment 4. In such case, participants can generate their responses at any point during a recall cycle, which makes too-late retrieval obsolete. The present finding that prolonged delay between study and test increases the hypermnesia effect aligns well with the retrieval practice hypothesis (Wallner & Bäuml, 2018). Retrieval practice is known to reduce forgetting, particularly when the recall task is demanding (Carpenter, 2011; Halamish & Bjork, 2011). Because retrieval after a prolonged delay should be more effortful than retrieval after a 0-min delay, item losses are expected to be reduced after prolonged delay, leading to an increase in the size of the hypermnesia effect, which is what the present results show.

The present findings are difficult to reconcile with two other prominent accounts of hypermnesia, the cumulative recall account (Roediger & Challis, 1989; Roediger et al., 1982) and the retrieval strategy account (Erdelyi & Becker, 1974; Mulligan, 2001). The cumulative recall account assumes that conditions yielding higher asymptotic recall should also yield greater hypermnesia. Because cumulative recall typically decreases with longer delay between study and test (see Experiments 1, 2a, and 2b), and, according to the account, asymptotic recall should be positively related to the size of the hypermnesia effect, one would expect the size of the hypermnesia effect to diminish, rather than increase, with prolonged delay. The retrieval strategy account faces difficulties in explaining why the hypermnesia effect was larger in the ILCR test constrained in timing and output order than in free recall, since, with the cued-recall format, opportunities for strategic variability should be minimal. Indeed, the implementation of a new or improved retrieval strategy typically modulates the output order in which items are produced at test (e.g., Chan, Manley, Davis, & Szpunar, 2018). Therefore, the use of an updated retrieval strategy should be disabled when the order of item retrieval is externally constrained, like in present Experiments 1-3. Consequently, this account would predict a smaller, rather than a larger, hypermnesia effect under the ILCR test constrained in timing and output order.

Further alternative explanations of the present findings warrant consideration.[1] First, an alternative to the "too-late retrieval" idea to explain the findings with the constrained ILCR format of Experiments 1-3 may be that, by depressing initial performance, the constrained ILCR format may impose an initial performance cost and the observed relatively large hypermnesia effects then reflect the progressive removal of that cost. For instance, retrieval cues may become more effective across recall cycles because items queried late in Test 1 show output interference (e.g., Roediger, 1973; Smith, 1971) but may no longer be subject to such effects when probed earlier on subsequent Tests 2 and 3, thereby yielding item gains. However, this explanation requires output interference effects not to persist across subsequent tests, whereas results from retrieval-induced forgetting, which is similar to output interference (e.g., Bäuml & Kliegl, 2025), suggest that output interference effects may last for 20 min or even hours (Anderson et al., 1994; Meixensperger & Bäuml, in press). A second alternative explanation is based on the observation that, across all experiments, hypermnesia was consistently larger in the condition that produced lower initial recall performance. Larger hypermnesia effects may thus arise whenever more room for item gains is left on subsequent tests. However, in all present experiments, there was ample room left for hypermnesia effects in the conditions with initially higher recall performance (e.g., free recall) to match or even outperform the hypermnesia effect observed with initially lower recall performance (e.g., ILCR). Finally, the finding of lower item losses in the 30-min relative to the 0-min delay conditions may reflect the fact that many weak items were already forgotten after the 30-min delay, whereas the remaining (mostly) strong items proved quite resistant to forgetting. Results from previous studies show that typically similar numbers of strong and weak items are forgotten over time (e.g., Rivera-Lares, Logie, Baddeley, & Della Sala, 2022; Slamecka & McElree, 1983), making it unlikely that mainly strong items survived the longer 30-min delay.

An a priori unexpected finding in Experiments 2a, 2b, and 3 may be that the constrained ILCR format produced not only larger item gains but also somewhat larger item losses across recall cycles compared to the free-recall format. Although the increase in item gains clearly outweighed the increase in item losses, thus producing greater hypermnesia overall, the question arises of why the cued-recall format should amplify item losses. One possibility is that the effects in item gains and item losses reflect the two sides of the same coin. If participants miss recall of early tested items during a recall cycle just to retrieve the items later in the cycle, such late retrieval may cause repeated covert retrieval of the items, just to increase chances that they will be accessible on the next recall cycle. However, as is suggested from research on retrieval induced forgetting (for reviews, see Bäuml & Kliegl, 2025; Storm & Levy, 2012), such covert retrieval may inhibit the memory representation of other (competing) list items and may thus reduce chances for previously recalled items to be recalled again on the next recall cycle. The increase in item losses with the cued-recall format may thus be a by-product of the increase in item gains.

Bäuml and Trißl's (2022) observation of a relatively large hypermnesia effect in a selective retrieval task arose with a constrained ILCR format at test and a short list of 15 study items. Using the same list length and the same delay interval, this study showed that Bäuml and Trißl's observation generalizes to a typical hypermnesia setup. In addition, the results also indicate that the possible contribution of list length to hypermnesia effects may be small. In fact, the pattern of results reported here with the free-recall format

[1] These explanations were suggested during the anonymous review process.

(Experiment 2b, Experiment 3) was very similar to the one reported in Wallner and Bäuml (2018), although Wallner and Bäuml used more typical long lists of 40 study items rather than the short lists of 15 study items used in the present study. While this parallel does not exclude that list length can have some influence on hypermnesia findings, it indicates that such effects may be rather limited in size.

One might argue that, in Experiments 1-3, item gains have been artificially inflated by the constrained test format and the observed increased hypermnesia effects with this recall format therefore may not constitute "real" hypermnesia effects. However, hypermnesia studies assess hypermnesia by comparing recall performance across multiple successive tests, and potential hypermnesia effects are regarded as genuine whenever recall performance increases across tests. This holds regardless of whether free or cued-recall formats are applied at test (e.g., Erdelyi & Becker, 1974; Otani & Hodge, 1991; Payne et al., 1993; Roediger & Thorpe, 1978) and is exactly how hypermnesia has also been addressed in the present study. Furthermore, test situations which constrain participants' recall by using an externally controlled output order and strict time limits for the recall of single memory contents have been widely employed in memory research (Anderson et al., 1994; Bäuml & Aslan, 2004). The present finding that recall performance may benefit from further recall cycles in such situations may be of relevance for these research areas, but it may also be of relevance for more applied situations, like, for instance, educational situations. A student in an oral examination may retrieve target information only after the examiner has already proceeded to the next question, and the students' knowledge of the correct response may thus become evident only if another recall attempt is provided. Examining whether such real-world cued-recall situations can produce hypermnesia effects similar to those observed in the present laboratory experiments may indeed be a high priority for future work on hypermnesia.

To conclude, the findings from the present study show that the magnitude of the hypermnesia effect can be increased when an ILCR format constrained in output order and timing rather than a free-recall format is used at test and can be enhanced even further with a prolonged delay between study and the recall tests. The effect of recall format was driven by an increase in item gains and the effect of delay by a decrease in item losses. This pattern of results is well captured by an explanation drawing on both the changes in cue set account of hypermnesia, which can account for the difference in item gains, and the retrieval practice hypothesis, which can account for the difference in item losses.

Funding Open Access funding enabled and organized by Projekt DEAL. Not applicable.

Availability of data and materials All study materials and data have been made publicly available on the Open Science Framework and can be found at https://osf.io/b6fyk/. None of the experiments was preregistered.

Code Availability Program code is not required to create stimuli or perform data analyses. Stimuli were either nouns which are documented in OSF (https://osf.io/b6fyk/) or simple line drawings which can be accessed via Severens et al. (2005) and Snodgrass and Vanderwart (1980). All significant data analyses can be performed on the basis of the information provided in the Results section of the manuscript.

Declarations

Conflicts of interest/Competing interests None.

Ethics approval The protocol employed in this study was deemed exempt by the ethical review board of Regensburg University. The experiments were carried out in accordance with the provisions of the World Medical Association 1964 Declaration of Helsinki.

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

Consent for publication Not applicable.

References

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

Aslan, A., & Bäuml, K.-H.T. (2007). Part-list cuing with and without item-specific probes: the role of encoding. *Psychonomic Bulletin & Review, 14*, 489–494. https://doi.org/10.3758/BF03194095

Bäuml, K.-H.T., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition, 32*, 610–617. https://doi.org/10.3758/BF03195852

Bäuml, K.-H.T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect-it enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1568–1581. https://doi.org/10.1037/xlm0000025

Bäuml, K.-H. T., & Kliegl, O. (2025). Retrieval-induced remembering and forgetting. In Wixted, J. T. (Ed.), *Cognitive psychology of memory, Vol. 4 of Learning and Memory: A comprehensive reference, 3rd edition*, pp. 311-327. Oxford: Academic Press. https://doi.org/10.1016/B978-0-443-15754-7.00003-1

Bäuml, K.-H.T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. Proceedings of the National Academy of Sciences (Vol. 119,. https://doi.org/10.1073/pnas.2114377119 Article e2114377119.

Belmore, S. M. (1981). Imagery and semantic elaboration in hypermnesia for words. *Journal of Experimental Psychology: Human Learning and Memory, 7*(3), 191–203. https://doi.org/10.1037/0278-7393.7.3.191

Bluck, S., Levine, L. J., & Laulhere, T. M. (1999). Autobiographical remembering and hypermnesia: A comparison of older and younger adults. *Psychology and Aging, 14*(4), 671–682. https://doi.org/10.1037/0882-7974.14.4.671

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1547–1552. https://doi.org/10.1037/a0024140

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268–276. https://doi.org/10.3758/BF03193405

Chan, J. C., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language, 102*, 83–96. https://doi.org/10.1016/j.jml.2018.05.007

Dunning, D., & Stern, L. B. (1992). Examining the generality of eyewitness hypermnesia: A close look at time delay and question type. *Applied Cognitive Psychology, 6*(7), 643–657. https://doi.org/10.1002/acp.2350060707

Erdelyi, M. H. (1996). *The recovery of unconscious memories: Hypermnesia and reminiscence*. Chicago: The University of Chicago Press.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*(1), 159–171. https://doi.org/10.1016/0010-0285(74)90008-5

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

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 801–812. https://doi.org/10.1037/a0023219

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10*(5), 562–567. https://doi.org/10.1016/S0022-5371(71)80029-4

Kelley, M. R., & Nairne, J. S. (2003). Remembering the forgotten? Reminiscence, hypermnesia and memory for order. *The Quarterly Journal of Experimental Psychology, 56*(4), 577–599. https://doi.org/10.1080/02724980244000530

Meixensperger, S. & Bäuml, K.-H. T. (in press). Items that are subject to retrieval-induced forgetting show slowed forgetting over time. *Memory*.

Montangero, J., Ivanyi, C. T., & de Saint-Hilaire, Z. (2003). Completeness and accuracy of morning reports after a recall cue: Comparison of dream and film reports. *Consciousness and Cognition, 12*(1), 49–62. https://doi.org/10.1016/S1053-8100(02)00029-6

Mulligan, N. W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 436–450. https://doi.org/10.1037/0278-7393.27.2.436

Mulligan, N. W. (2006). Hypermnesia and total retrieval time. *Memory, 14*, 502–518. https://doi.org/10.1080/09658210500513438

Otani, H., & Griffith, J. D. (1998). Hypermnesia for prose. *The Journal of General Psychology, 125*(2), 147–155. https://doi.org/10.1080/00221309809595539

Otani, H., & Hodge, M. H. (1991). Does hypermnesia occur in recognition and cued recall? *The American Journal of Psychology, 104*(1), 101–116. https://doi.org/10.2307/1422853

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101*(1), 5–27. https://doi.org/10.1037/0033-2909.101.1.5

Payne, D. G., Hembrooke, H. A., & Anastasi, J. S. (1993). Hypermnesia in free recall and cued recall. *Memory & Cognition, 21*(1), 48–62. https://doi.org/10.3758/BF03211164

Raaijmakers, J. G., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In B. Ross (Ed.), *The Psychology of Learning and Motivation.* (Vol. 14). Academic Press: Elsevier Inc.

Rivera-Lares, K., Logie, R., Baddeley, A., & Della Sala, S. (2022). Rate of forgetting is independent of initial degree of learning. *Memory & Cognition, 50*(8), 1706–1718. https://doi.org/10.3758/s13421-021-01271-1

Roediger, H. L., III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 644–657. https://doi.org/10.1016/S0022-5371(73)80044-1

Roediger, H. L., & Challis, B. H. I. (1989). Hypermnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.), *Current Issues in Cognitive Processes: The Tulane Floweree Symposium on Cognition Hillside.* NJ: Erlbaum.

Roediger, H. L. I., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693

Roediger, H. L., & Payne, D. G. I. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(1), 66–72. https://doi.org/10.1037/0278-7393.8.1.66

Roediger, H. L., & Thorpe, L. A. I. (1978). The role of recall time in producing hypermnesia. *Memory and Cognition, 6*, 296–305. https://doi.org/10.3758/BF03197459

Roediger, H. L., Payne, D. G. I., Gillespie, G. L., & Lean, D. S. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning and Verbal Behavior, 21*(6), 635–655. https://doi.org/10.1016/S0022-5371(82)90810-6

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(3), 384. https://doi.org/10.1037/0278-7393.9.3.384

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*(4), 400–408. https://doi.org/10.1016/S0022-5371(71)80039-7

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–214. https://doi.org/10.1037/0278-7393.6.2.174

Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition, 40*(6), 827–843. https://doi.org/10.3758/s13421-012-0211-7

Wallner, L., & Bäuml, K.-H.T. (2018). Hypermnesia and the role of delay between study and test. *Memory and Cognition, 46*, 878–894. https://doi.org/10.3758/s13421-018-0809-5

Wheeler, M. A., & Roediger, H. L. I. (1985). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science,3*(4), 240–245. https://doi.org/10.1111/j.1467-9280.1992.tb00036.x

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