



Hypermnesia and the Role of Delay between Study and Test

Lisa A. Wallner¹ · Karl-Heinz T. Bäuml¹

Published online: 25 April 2018
© Psychonomic Society, Inc. 2018

Abstract

Hypermnesia is increased recall across repeated tests in the absence of any further study opportunities. Although over the years many factors have been identified that influence hypermnesia, to date not much is known about the role of delay between study and test for the effect. This study addressed the issue in four experiments. Employing both words and pictures as study material, we compared hypermnesia after shorter delay (3 min or 11.5 min) and longer delay (24 h or 1 week) between study and test. Recall occurred over three successive tests, using both free recall (Experiments 1, 2, and 4) and forced recall testing (Experiment 3). In forced recall tests, subjects are instructed to recall as many items as possible, but if unable to remember all studied items, to fill in the remaining spaces with their best guesses. With free recall testing, hypermnesia increased with delay and the effect was driven mainly by reduced item losses between tests. These results suggest a link between hypermnesia and the testing effect, which shows that demanding retrieval practice, as it happens after longer delay, can improve recall by reducing the forgetting of the practiced items. In contrast, with forced recall testing, hypermnesia decreased with delay and was even absent after longer delay. The findings indicate that recall format can influence hypermnesia and different mechanisms may mediate the effects of repeated testing in the two recall conditions.

Keywords Hypermnesia · Multiple tests · Retrieval practice · Delay · Testing effect

Hypermnesia refers to enhanced recall with repeated recall attempts. In a typical experiment on hypermnesia, participants study a set of items, like words or pictures, and are then presented with a series of successive recall tests, in each of which they are asked to recall the previously studied items. Across tests, some items are recalled on later tests that were not recalled in prior tests (item gains), whereas other items recalled on prior tests are not recalled on a later test (item losses). Hypermnesia arises if item gains exceed item losses, and consequently, a net increase in the number of items recalled across tests results. In contrast, net forgetting is generated if item losses exceed item gains (for reviews, see Erdelyi, 1996; Payne, 1987).

Hypermnesia is a robust effect that was demonstrated in quite different experimental settings. It was shown in a variety of list-learning experiments, for instance, employing

unrelated words, associated word pairs, pictures, foreign language vocabulary, or nonsense syllables (e.g., Belmore, 1981; Kelley & Nairne, 2003; Mulligan, 2001; Roediger & Payne, 1982). It arose with prose passages (Otani & Griffith, 1998; Wheeler & Roediger, 1985) and films (Montangero et al., 2003), and was demonstrated in studies on eyewitness memory (Dunning & Stern, 1992) and autobiographical memory (Bluck et al., 1999).

Accounts of hypermnesia

Despite the large number of studies that has been conducted on hypermnesia, to date it is still unclear exactly which mechanisms mediate the effect. Over the years, a variety of different explanations for the effect have emerged. One of the most prominent explanations is the *cumulative recall hypothesis* (Roediger & Challis, 1989; Roediger et al., 1982). This hypothesis assumes that hypermnesia is a function of the cumulative level of recall of items and that study conditions producing high levels of recall are more likely to exhibit hypermnesia than study conditions producing lower levels of recall. In this approach, the end of the first recall test, which typically lasts between 5 and 7

✉ Karl-Heinz T. Bäuml
karl-heinz.bauml@ur.de
Lisa A. Wallner
lisa.wallner@ur.de

¹ Department of Experimental Psychology, Regensburg University, 93053 Regensburg, Germany

min, is considered an interruption of recall. Thus, if an experimental condition has not yet reached its asymptotic recall level at the end of this test—i.e., the level that could be produced given unlimited recall time—then the additional retrieval time afforded by the subsequent test can produce item gains.

Another account of hypermnesia is the *changes in cue set hypothesis* (Raaijmakers & Shiffrin, 1980; Roediger & Thorpe, 1978). This hypothesis suggests that the cue set that people use to sample and recover memories over longer test intervals can change depending on the items “sampled” as retrieval cues. Because new cue sets arise with newly recalled information, on a later test, alternative retrieval routes may be used, which may lead to retrieval of previously unrecalled information and thus improve recall performance. Yet another account of hypermnesia is the *retrieval strategy hypothesis*, which explains hypermnesia by improved retrieval strategies and enhanced organization arising from retrieval practice in repeated testing (Erdelyi & Becker, 1974; Mulligan, 2001). According to this view, accessibility of information on a later test may be greater than that on an earlier test, because the earlier test permits more efficient organization of recalled material, so that, on the later test, the already recalled material can be retrieved again more quickly, with time remaining for the recall of new material. More organized retrieval strategies may also limit the number of item losses between tests, thus further increasing net recall levels (McDaniel et al., 1998).

Although each of the accounts can explain important findings in the hypermnesia literature, none of them can account for the full range of experimental results. For instance, while the cumulative recall hypothesis can explain the positive relation between variables affecting recall levels (e.g., imagery, semantic elaboration) and the magnitude of hypermnesia (Roediger & Challis, 1989; Roediger et al., 1982), the functional equivalence between single and repeated recall tests of equal total duration, which is predicted by the hypothesis, has repeatedly been challenged (Mulligan, 2005, 2006). Similarly, while the retrieval strategy hypothesis can account for the fact that retrieval strategies become increasingly organized over multiple recall tests and appear to contribute to hypermnesia (McDaniel et al., 1998; Mulligan, 2001), the hypothesis, for instance, has trouble explaining the picture–word difference, the very robust finding of higher hypermnesia for pictures than words (Payne, 1987). Finally, the changes in cue set hypothesis can describe several basic findings in the hypermnesia literature (Raaijmakers & Shiffrin, 1980), however, more direct tests of the hypothesis are rare. Moreover, like the cumulative recall hypothesis, this hypothesis focuses on item gains and is largely silent on item losses that may occur across subsequent recall tests.

The possible role of delay between study and test for hypermnesia

A factor that can speak to these accounts of hypermnesia is the role of delay between study and test. In a typical experiment on hypermnesia, the initial test occurs shortly after study without any major delay between study and test. Indeed, most hypermnesia studies employed a short delay between study and test of 1 or 2 min only, mainly to distribute the recall protocols or give detailed test instructions (e.g., Bergstein & Erdelyi, 2008; Kelley & Nairne, 2003; Mulligan, 2002; Payne & Roediger, 1987). Other studies additionally included filler tasks of 2 or 3 min to reduce possible recency effects (e.g., Mulligan, 2005; Otani, Widner, Whiteman, & Louis, 1999), employed a delay of 5 min with the subjects’ instruction to think silently about the list items (Shapiro & Erdelyi, 1974), or employed a delay of about 12 min, asking subjects to participate in a distractor task and complete a questionnaire (Wheeler & Roediger, 1985). For this range of relatively short retention intervals, there is no indication yet that delay influences hypermnesia.

However, longer retention intervals may well influence hypermnesia. On the basis of the cumulative recall hypothesis, for instance, one may expect that hypermnesia decreases with an increase in delay between study and test. Indeed, because longer delays generally reduce (cumulative) recall levels, and, according to the hypothesis, recall levels are positively related to the magnitude of hypermnesia, hypermnesia should be smaller after longer than shorter retention intervals and item gains decrease with delay. In contrast, on the basis of the changes in cue set hypothesis, one may expect that hypermnesia increases with delay. Delay causes context shift (e.g., Bower, 1972; Estes, 1955) and, after context shift, retrieval of some first items can reactivate the study context and facilitate recall of the other items (e.g., Bäuml & Schlichting, 2014; Howard & Kahana, 1999; Wallner & Bäuml, 2017). If such context reactivation was not yet complete at the end of the first recall test but extended to subsequent tests, then a longer delay between study and test may lead to more extensive changes in cue set across tests than a shorter delay, and thus enhance item gains and increase hypermnesia. Likewise, on the basis of the retrieval strategy hypothesis, one may also expect enhanced hypermnesia after longer delay. If delayed recall led to more organized retrieval strategies compared to recall after shorter delay, for instance, because recall after delay can be more challenging, then according to the hypothesis, repeated testing after longer delay may both enhance item gains and reduce item losses.

Changes in cue set and improved retrieval strategies after delay may not be the only reasons to expect increased hypermnesia after prolonged retention intervals.

Differences in retrieval practice effects after short versus long delay may also influence hypermnesia, a view referred to as the *retrieval practice hypothesis* in the following. In fact, from the testing effect literature, it is well known that (i) prior retrieval makes practiced items more accessible on subsequent tests and reduces the forgetting of the items (e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006), and (ii) such beneficial effects of retrieval practice are particularly strong if retrieval practice is demanding, like, for instance, in the presence of weak retrieval cues or in the presence of interference (e.g., Bäuml, Holterman, & Abel, 2014; Carpenter, 2011; Halamish & Bjork, 2011; Pyc & Rawson, 2009). Because, in general, longer delay should also make retrieval more demanding, the findings from the testing effect literature suggest that, after longer delay, retrieval on an initial test may increase hypermnesia by reducing the forgetting of the initially recalled items.

While the changes in cue set, the improved retrieval strategies, and the retrieval practice hypotheses lead to the expectation of increased hypermnesia after delay, the three hypotheses differ in their expectations on item gains and item losses. Because the changes in cue set hypothesis are primarily framed around item gains, it suggests increased item gains with delay, without making detailed suggestions regarding item losses; the retrieval strategy hypothesis leads to the expectation of both enhanced item gains and reduced item losses with delay; and the retrieval practice hypothesis suggests mainly a reduction in item losses with delay. Table 1 provides an overview of these expectations.

Prior work on the role of delay for hypermnesia

To the best of our knowledge, there are only three studies in the literature yet that employed retention intervals of more than 12 min between study and test to examine the role of delay for hypermnesia. In one study, Dunning and Stern

Table 1 Overview of expectations from single accounts of hypermnesia regarding the effects of increased delay between study and test on net recall, item gains, and item losses

	Net recall	Item gains	Item losses
CR	↓	↓	–
CCS	↑	↑	–
RS	↑	↑	↓
RP	↑	0	↓

Note. CR = cumulative recall hypothesis; CCS = changes in cue set hypothesis; RS = retrieval strategy hypothesis; RP = retrieval practice hypothesis. Effects may be expected to increase (↑), decrease (↓), or remain constant (0). Alternatively, there may be no expectation (–) by the particular account

(1992; Experiment 2) investigated whether hypermnesia in eyewitness memory depends on delay between study and test. Subjects viewed videotapes reenacting several types of crimes and, after varying delay, were asked to provide accounts of the incident on three successive free recall tests. The initial interview occurred immediately after watching the video tapes, after a 3-day delay, or a 1-week delay. Results revealed typical time-dependent forgetting for the number of correctly recalled facts with increasing delay. Above all, they showed hypermnesia without any influence of delay on the size of the effect. In the second study, Roediger and Payne (1982) presented subjects a list of pictures and then gave them three successive free recall tests. The first test was presented immediately after study, or was delayed by reading a prose passage by 18 min. Similar to Dunning and Stern (1992), the results showed hypermnesia, but again there was no effect of delay on the size of the effect. In the third study, Wheeler and Roediger (1985; Experiment 1) examined a number of factors of possible relevance for effects of repeated testing, but a subset of the experimental conditions is directly related to the present study. In this subset, subjects studied a list of pictures, either together with their names or embedded in a story, and, after study, received three immediate tests (the 3-3 condition) or three tests after a 1-week delay (the 0-3 condition). Results revealed typical time-dependent forgetting. More important, they showed hypermnesia after the short delay but no hypermnesia after the prolonged delay.

The results of the three studies thus provide mixed results. The results of two studies indicate that hypermnesia is not modulated by delay between study and test, whereas the results of the third study suggest decreased, if not absent, hypermnesia after prolonged retention interval. There are several factors that may be responsible for these results. For instance, Dunning and Stern (1992) employed a very small sample of subjects, with 8-11 subjects only in each single delay condition, a sample that may have been too low in size to detect significant influences of delay on hypermnesia. In Roediger and Payne (1982), the delay manipulation did not induce any time-dependent forgetting, which indicates that the manipulation may have been ineffective and thus have limited the room for influences of delay on hypermnesia. In contrast to Dunning and Stern (1992) and Roediger and Payne (1982), who used free recall at test, Wheeler and Roediger (1985) employed a forced recall format. In this format, subjects are given recall sheets with a separate line for each single to-be-recalled item at test; subjects are instructed to recall as many items as possible, but if unable to remember all studied items, to fill in the remaining spaces with their best guesses.¹ Roediger and Payne (1985)

¹Forced recall was introduced into the hypermnesia literature by Erdelyi and Becker (1974) with the goal to hold response criteria, i.e., recall threshold, constant across tests.

provided evidence that free and forced recall can lead to similar hypermnesia effects, but whether this pattern, which arose for short delay conditions, generalizes to prolonged retention interval is unclear (see below). To clarify the role of delay between study and test for hypermnesia, fresh experiments are necessary that (i) include a sufficiently large sample of subjects, (ii) employ delay conditions that induce robust time-dependent forgetting, (iii) examine the possible influence of recall format (free versus forced recall) on hypermnesia after longer delay, and (iv) include not only an analysis of net recall but also of item gains and item losses. In fact, in none of the three previous studies item gains and item losses were analyzed.

The present study

The present study reports the results of four experiments designed to examine whether, and if so, how delay between study and test influences hypermnesia. In each experiment, subjects studied a list of items and, after a delay, were repeatedly asked to recall the previously studied material. Critically, in all four experiments, delay between study and the first recall test was manipulated, using either a short retention interval of 3 min (Experiments 1 and 2) or 11.5 min (Experiments 3 and 4), or a prolonged retention interval of 24 h (Experiments 1 and 2) or 1 week (Experiments 3 and 4). In Experiment 1, subjects at study rated a list of unrelated words to be living or nonliving, whereas in Experiments 2, 3, and 4, they studied a list of pictures. At test, subjects in all four experiments participated in a series of three successive recall tests. In Experiments 1, 2, and 4, the recall format was free recall, in Experiment 3 it was forced recall. In all experiments, we expected to replicate typical hypermnesia in the short delay condition with an increase in net recall across tests, irrespective of recall format. The critical question was whether delay would influence this beneficial effect, and if so, whether it reduced or enhanced hypermnesia. The results of the experiments will improve our understanding of the role of delay for hypermnesia and provide new information on the mechanisms contributing to the effect.

Experiment 1

Experiment 1 examined the effect of delay between study and initial test on hypermnesia employing lists of unrelated words. Subjects were presented the words and, for each single word, were asked to indicate if it was living or nonliving (e.g., Belmore, 1981). After a delay of 3 min or 24 h, subjects participated in three successive free recall tests, in each of which they were asked to remember and write down as many of the previously rated items

as possible, independent of what they had remembered in possible preceding tests. On the basis of the cumulative recall hypothesis, one may expect larger hypermnesia after the short than the long retention interval, which would be consistent with Wheeler and Roediger's (1992) finding. In contrast, on the basis of the changes in cue set, the retrieval strategy, and the retrieval practice hypotheses, one may expect larger hypermnesia after the long than the short delay. Following the changes in cue set hypothesis, such increased hypermnesia may be mediated mainly by enhanced item gains, whereas following the retrieval practice hypothesis it may be mediated mainly by reduced item losses. Following the retrieval strategy hypothesis, both effects may arise.

Method

Participants To ensure that a possible effect of delay on hypermnesia could be detected in the present experiment, an analysis of test power was conducted with the G*Power program (version 3, Faul, Erdfelder, Lang, & Buchner, 2007) to estimate the number of participants required. This analysis revealed that, to detect a small-to-medium sized effect ($f=0.20$; Cohen, 1988) for the critical interaction with a probability of $1-\beta=.80$ and $\alpha=.05$, 42 participants were required. Following this analysis, 42 students of Regensburg University took part in the experiment ($M=22.19$ years, range 18–30 years, 64.3% female). All participants spoke German as native language and took part on a voluntary basis. They received monetary reward or course credit for their participation.

Materials For counterbalancing purposes, two study lists (A, B) were constructed, each containing 48 labels of line-drawing pictures selected from the Snodgrass and Vanderwart (1980) norms (see Appendix A). All items were high-imagery nouns. Thirty percent of the items were selected as “living” and the rest as “nonliving”. Items were chosen that elicited very high name agreement (98–100% according to the Snodgrass & Vanderwart norms) and had single word names. Two of the 48 items of a list served as primacy and two other items as recency items in this experiment. The remaining 44 items served as target items (see also Mulligan, 2006). All items were translated into German.

Design The experiment had a 2×3 repeated measures design with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). Participants were tested on a study list 3 min after study (short delay) and after a delay of 24 h (long delay). At test, in both delay conditions subjects recalled the studied items in three successive free recall tests, which were separated by short distractor tasks. Assignment of conditions and lists was counterbalanced.

Procedure Each participant completed two experimental blocks in counterbalanced order, one in the short and one in the long delay condition. The blocks were separated by a 5 min break, in which subjects played Tetris. Prior to the study phase of each block, participants were informed that they would see a list of words and that they should try to rate the words whether they were “living” or “nonliving” (see Belmore, 1981). All words were presented individually on a screen for 5 s each and in random order. The entire list was presented twice in immediate succession (e.g., Mulligan, 2006). In the short delay condition, subjects were then asked to count backwards from a three-digit number for 3 min, while in the long delay condition, subjects were disbanded at this point and were asked to come back at the same time the next day. The test phase was identical for the two delay conditions. Participants completed three successive free recall tests, each lasting for 5 min. At the beginning of each test, a blank sheet was distributed with the instruction to report as many of the previously studied items as possible, independent of what they may have remembered in possible preceding tests. Between the tests, participants solved arithmetic problems for 3 min.

Results

Separately for the short delay (3 min) and the long delay (24 h) conditions, Table 2 shows (i) net recall, i.e., number of correctly recalled words on each single test, (ii) item gains and item losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusion rates, i.e., number of recalled items not presented during study of the list.

Table 2 Net recall, item gains, item losses, and intrusions in Experiment 1, separately for the short delay (3 min) and the long delay (24 h) condition

Words	3-min delay		24-h delay	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Free recall				
Net recall				
Test 1	22.02	8.89	14.54	8.49
Test 2	21.95	9.13	15.58	9.29
Test 3	22.29	9.14	16.89	9.64
Gains and losses				
Gains 1 - 2	1.10	1.53	1.43	1.81
Gains 2 - 3	.88	1.04	1.43	1.86
Losses 1 - 2	1.17	1.46	.38	.73
Losses 2 - 3	.55	.94	.10	.30
Intrusions				
Test 1	.38	.82	.62	1.61
Test 2	.40	.86	1.14	2.18
Test 3	.55	1.13	1.26	2.06

Net recall A 2×3 analysis of variance (ANOVA) with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3) showed a main effect of DELAY, $F(1, 41) = 22.45$, $MSE = 115.49$, $p < .001$, $\eta^2 = .35$, which demonstrates typical time-dependent forgetting. It also revealed a main effect of TEST, $F(2, 82) = 12.69$, $MSE = 8.89$, $p < .001$, $\eta^2 = .24$, indicating hypermnesia. More important, there was a significant interaction between the two factors, $F(2, 82) = 11.36$, $MSE = 2.04$, $p < .001$, $\eta^2 = .22$, suggesting that the amount of increase in net recall across tests varied with delay. Consistently, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed no significant main effect of TEST after 3 min, $F(2, 82) < 1$, but a significant main effect of TEST after 24 h, $F(2, 82) = 20.67$, $MSE = 2.82$, $p < .001$, $\eta^2 = .34$, suggesting that hypermnesia arose after the long but not the short delay. After 24 h, recall on the second test exceeded that on the first test, $t(41) = 3.50$, $p = .001$, $d = .12$, and recall on the third test exceeded that on the second test, $t(41) = 4.23$, $p < .001$, $d = .14$.²

Item gains and item losses We next analyzed item gains and item losses across tests. Gains on the second test were studied items reported on the second test but not on the first test, and gains on the third test were items reported on the third test but not on the second test. Likewise, losses on the second test were items reported on the first test but not the second, and losses on the third test were items reported on the second test but not the third. Regarding item gains, a 2×2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed no main effect of DELAY, $F(1, 41) = 3.53$, $MSE = 2.31$, $p = .067$, $\eta^2 = .08$, no main effect of TEST, $F(1, 41) < 1$, and no interaction between the two factors, $F(1, 41) < 1$. Regarding item losses, the same ANOVA showed a main effect of DELAY, $F(1, 41) = 21.01$, $MSE = .77$, $p < .001$, $\eta^2 = .34$, as well as a main effect of TEST, $F(1, 41) = 8.62$, $MSE = .10$, $p = .005$, $\eta^2 = .17$, with more losses in the short delay condition than in the long delay condition, and more losses between the first and the second test than between the second and the third test. The interaction was not significant, $F(1, 41) = 1.35$, $MSE = .86$, $p = .215$, $\eta^2 = .03$.

Intrusions Analysis of intrusions may provide information on whether response criteria change across tests and delay conditions. Intrusions were analyzed with a 2×3 ANOVA with the within-subjects factors of DELAY (long, short) and TEST (test 1, test 2, test 3). It revealed significant main effects of DELAY, $F(1, 41) = 4.35$, $MSE = 4.60$,

²Calculation of effect sizes d followed Cohen's (1977) d statistics.

$p = .043$, $\eta^2 = .10$, and TEST, $F(2, 82) = 10.07$, $MSE = .36$, $p < .001$, $\eta^2 = .20$, showing that there were more intrusions after 24 h than after 3 min, and that intrusions increased across tests. There was also a significant interaction between the two factors, $F(2, 82) = 4.95$, $MSE = .38$, $p = .009$, $\eta^2 = .11$, suggesting that delay enhances the increase in intrusions with repeated testing.³

Discussion

Results show an increase of net recall across tests reflecting typical hypermnnesia. This increase, however, varied with the delay between study and test. Hypermnnesia was larger after the long than the short delay and was even nonsignificant in the short delay condition. Moreover, the increase in hypermnnesia with delay was primarily driven by reduced item losses across tests and was hardly affected by enhanced item gains. The findings on net recall are inconsistent with the cumulative recall hypothesis, which predicts reduced hypermnnesia with prolongation of delay, but are consistent with the changes in cue set, the retrieval strategy, and the retrieval practice hypotheses. The finding that the effect is mainly due to a reduction in item losses but less, if at all, to enhanced item gains favors the retrieval practice hypothesis over the other two accounts (compare Table 1). Intrusions increased across tests and with delay, which points to changes in response criteria. It is unlikely that changes in response criteria mediated the effect of delay on hypermnnesia in the present experiment, however. In fact, loosening the criterion with delay should increase item gains more than affecting item losses, which is not what the present results show. Before drawing more firm conclusions on the issue, it is the goal of Experiment 2 to replicate the present pattern of results.

Experiment 2

A factor critically contributing to hypermnnesia is stimulus material. Since Ballard's (1913) demonstration of the role of stimulus material for hypermnnesia, many studies showed that hypermnnesia effects arise fairly easily with some kind of stimulus material (e.g., pictures; Erdelyi & Kleinbard, 1978; Madigan, 1976; Madigan & Lawrence, 1980), but may be harder to get with others (e.g., lists of unrelated words, (Nelson & MacLeod, 1974; Tulving, 1967;

Wilkinson & Koestler, 1983). In his review, Payne (1987) integrated 172 studies, and summarized that 96% of the experiments using simple pictures produced hypermnnesia, whereas only 46% of the experiments using word lists did. Hence, the finding of nonsignificant hypermnnesia with words in the short delay condition of Experiment 1 is not atypical in research on hypermnnesia. Because hypermnnesia is more readily found when pictures are used as study material and because words and pictures sometimes produce different results regarding hypermnnesia (e.g., Erdelyi & Becker, 1974; Payne, 1986), we aimed to repeat Experiment 1 with pictures as study material. We presented the same set of items in Experiment 2 as in Experiment 1, but showed the items' pictorial representations in the study phase. Doing so, we expected reliable hypermnnesia in the short delay condition. The critical question then was if hypermnnesia was again increased in the prolonged retention interval condition and whether such increase in net recall was again mainly driven by reduced item losses.

Method

Participants Another 42 students of Regensburg University took part in the experiment ($M = 22.14$ years, range, 17–32 years, 64.3% female). All participants spoke German as native language and took part on a voluntary basis. Again, they received monetary reward or course credit for their participation.

Materials We employed the same two study lists (A, B) as in Experiment 1. However, in contrast to Experiment 1, not the labels of the pictures were presented in the study phase, but the line-drawings themselves (see Snodgrass & Vanderwart, 1980). As in Experiment 1, the same four buffer items of each list were applied to control for primacy and recency effects.

Design and procedure Design and procedure were identical to Experiment 1, with the only exception that participants in the study phase were not instructed to rate the words to be "living" or "nonliving". Rather, participants were informed that they would see a list of pictures and that they should try to remember them for a later memory test (e.g., Mulligan, 2006).

Results

Separately for the short delay (3 min) and the long delay (24 h) conditions, Table 3 shows (i) net recall, i.e., number of correctly recalled pictures on each single test, (ii) item gains and item losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusions on each single recall test.

³Number of intrusions was fairly low in this experiment. We therefore did not differ between inter-list intrusions (words from block 1 intruding during recall in block 2) and intrusions caused by items not presented in any of the two experimental blocks. The same holds for Experiments 2 and 4.

Table 3 Net recall, item gains, item losses, and intrusions in Experiment 2, separately for the short delay (3 min) and the long delay (24 h) condition

Pictures	3-min delay		24-h delay	
	<i>M</i>	SD	<i>M</i>	SD
Free recall				
Net recall				
Test 1	27.57	7.60	19.02	8.64
Test 2	27.74	8.27	20.12	8.76
Test 3	28.71	8.31	21.00	9.08
Gains and losses				
Gains 1 - 2	1.38	1.50	1.52	2.18
Gains 2 - 3	1.86	2.18	1.12	1.27
Losses 1 - 2	1.21	1.57	.43	.74
Losses 2 - 3	.74	1.36	.26	.63
Intrusions				
Test 1	.45	.59	1.05	2.27
Test 2	.48	.67	1.19	2.88
Test 3	.57	.77	1.26	2.80

Net recall The net recall data were scored using a conservative scoring method, in which the recalled name had to match the German translation of the picture name given by the Snodgrass and Vanderwart (1980) norms.⁴ The net recall data were analyzed with a 2×3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). There was a main effect of DELAY, $F(1, 41) = 57.07$, $MSE = 69.97$, $p < .001$, $\eta^2 = .58$, showing typical time-dependent forgetting, and a main effect of TEST, $F(2, 82) = 23.12$, $MSE = 2.24$, $p < .001$, $\eta^2 = .36$, indicating increased recall across tests, i.e., hypermnnesia. In addition, there was a significant interaction between the two factors, $F(2, 82) = 3.20$, $MSE = 1.71$, $p = .046$, $\eta^2 = .07$, suggesting that the test-induced increase in recall varied with delay condition. This held while there was significant hypermnnesia in both delay conditions. In fact, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed a significant main effect of test in both the short delay condition, $F(2, 82) = 8.63$, $MSE = 1.85$, $p < .001$, $\eta^2 = .17$, and the long delay condition, $F(2, 82) = 19.67$, $MSE = 2.09$, $p < .001$, $\eta^2 = .32$. In the short delay condition, recall on the first and the second tests did not differ significantly, $t(41) < 1$, but recall on the second and third tests did,

$t(41) = 3.65$, $p = .001$, $d = .12$. In contrast, in the long delay condition both recall on the second test exceeded that on the first test, $t(41) = 3.10$, $p = .004$, $d = .13$, and recall on the third test exceeded that on the second, $t(41) = 4.03$, $p < .001$, $d = .10$.

Item gains and item losses Regarding item gains, a 2×2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed no main effect of DELAY, no main effect of TEST, and no interaction between the factors, all $F's(1, 41) < 2.25$, $MSE's < 3.62$, $p's > .141$, $\eta's^2 < .05$. The same ANOVA for item losses showed a significant main effect of DELAY, $F(1, 41) = 12.24$, $MSE = 1.37$, $p = .001$, $\eta^2 = .23$, suggesting that item losses in the short delay condition exceeded item losses in the long delay condition. There was a significant main effect of TEST, $F(1, 41) = 4.40$, $MSE = .99$, $p = .042$, $\eta^2 = .10$, indicating that item losses significantly decreased across tests, with more losses between the first and the second test than between the second and the third test. The interaction was not significant, $F(1, 41) = 1.04$, $MSE = .97$, $p = .314$, $\eta^2 = .03$.

Intrusions Intrusions were analyzed with a 2×3 ANOVA with the within-subjects factors of DELAY (3 min, 24 h) and TEST (test 1, test 2, test 3). It revealed no main effect of DELAY, $F(1, 41) = 3.03$, $MSE = 9.23$, $p = .089$, $\eta^2 < .07$, no main effect of TEST, $F(2, 82) = 2.50$, $MSE = .23$, $p = .089$, $\eta^2 < .06$, and no interaction between the two factors, $F(2, 82) < 1$.

Discussion

Using pictures as stimulus material, the results of this experiment showed expected hypermnnesia in the short delay condition. More important, like in Experiment 1, hypermnnesia was influenced by the delay between study and test and was larger after the longer than the shorter delay. Also like in Experiment 1, this effect of delay was mainly driven by a reduction in item losses across tests in the long delay condition. There were no effects regarding intrusions, suggesting that, in this experiment, response criteria were roughly constant. The observed increase in net recall with delay is again consistent with the changes in cue set, the retrieval strategy, and the retrieval practice hypotheses, although the observed reduction in item losses favors the retrieval practice explanation of the present results.

Additional analyses

In contrast to the between-subjects design employed in the three extant studies on the issue (Dunning & Stern, 1992;

⁴There was an alternative, more liberal scoring method, in which the experimenter decided if the recalled name corresponded to a picture from the study list, even if the label was a variant of the normed name (e.g., Mulligan, 2006). The two scoring methods produced equivalent results, thus, only the data from the conservative scoring method are reported here. The not matched labels of recalled pictures, e.g., crab instead of lobster, were added to the number of intrusions.

Table 4 Net recall, item gains, item losses, and intrusions pooled over the first experimental blocks of Experiments 1 and 2. Results are shown separately for the short delay (3 min) and the long delay (24 h) condition

Words + pictures	3-min delay		24-h delay	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Free recall				
Net recall				
Test 1	24.38	8.34	16.27	7.98
Test 2	24.33	9.06	17.20	8.37
Test 3	25.05	9.06	17.98	8.55
Mean number of gains and losses				
Gains 1 - 2	1.17	1.40	1.33	1.78
Gains 2 - 3	1.29	1.52	1.00	1.25
Losses 1 - 2	1.21	1.44	.40	.77
Losses 2 - 3	.45	.94	.24	.58
Intrusions				
Test 1	.48	.80	1.14	2.61
Test 2	.40	.83	1.64	3.46
Test 3	.52	1.04	1.76	3.21

Roediger & Payne, 1982; Wheeler & Roediger, 1985), in Experiments 1 and 2, each subject participated in both the short delay and the long delay conditions. Because this feature may have created order effects, we reanalyzed the data of the two experiments, this time including each subject's first block data only into the analysis. To maintain sufficient statistical power (see Methods of Experiment 1 above), we pooled the data of the two experiments to get again 42 participants in each delay condition. Table 4 shows net recall, item gains, item losses, and intrusions for the pooled data.

Statistical analysis of the pooled data replicated the main results for the two single experiments. Regarding net recall, a 2×3 ANOVA with the within-subjects factor of TEST (test 1, test 2, test 3) and the between-subjects factor of DELAY (short, long) showed a main effect of DELAY, $F(1, 82) = 16.13$, $MSE = 216.04$, $p < .001$, $\eta^2 = .16$, a main effect of TEST, $F(2, 164) = 14.66$, $MSE = 2.08$, $p < .001$, $\eta^2 = .15$, and a significant interaction between the two factors, $F(2, 164) = 3.46$, $MSE = 2.08$, $p = .034$, $\eta^2 = .04$. Recall increased across tests in the long delay condition, $F(2, 82) = 15.09$, $MSE = 2.05$, $p < .001$, $\eta^2 = .27$, and in the short delay condition, $F(2, 82) = 3.19$, $MSE = 2.10$, $p = .046$, $\eta^2 = .07$. In the long delay condition, recall on the second test exceeded that on the first test, $t(41) = 3.18$, $p = .003$, $d = .11$, and recall on the third test exceeded that on the second, $t(41) = 3.50$, $p = .001$, $d = .09$. In the short delay condition, recall on the second test did not differ to that on the first test, $t(41) < 1$, but recall on the third test exceeded that on the second, $t(41) = 3.06$, $p = .004$, $d = .08$.

Regarding item gains, a 2×2 ANOVA with the between-subjects factor of DELAY (short, long) and the within-subjects factor of TEST (test 2, test 3) revealed no main effects, both F 's(1, 82) < 1 , and no interaction between the two factors, $F(1, 82) = 1.26$, $MSE = 1.72$, $p = .267$, $\eta^2 = .02$. The same ANOVA for item losses showed a significant main effect of DELAY, $F(1, 82) = 8.98$, $MSE = 1.23$, $p = .004$, $\eta^2 = .10$, and a significant main effect of TEST, $F(1, 82) = 12.64$, $MSE = .72$, $p < .001$, $\eta^2 = .13$, indicating that item losses in the short delay condition exceeded item losses in the long delay condition and that there were more losses between test 1 and test 2 than between test 2 and test 3. There was also a significant interaction between the two factors, $F(1, 82) = 5.20$, $MSE = .72$, $p = .025$, $\eta^2 = .06$, suggesting that the reduction in item losses in the long delay condition was present mainly from the first to the second recall test. At least numerically, this same interaction was also present in the two single experiments reported above.

Regarding intrusions, a 2×3 ANOVA with the between-subjects factors of DELAY (long, short) and the within-subjects factor of TEST (test 1, test 2, test 3) showed significant main effects of DELAY, $F(1, 82) = 4.62$, $MSE = 14.98$, $p = .035$, $\eta^2 = .05$, and TEST, $F(2, 164) = 6.41$, $MSE = .37$, $p = .002$, $\eta^2 = .07$, suggesting that there were more intrusions after a long delay and that intrusions raised across tests. As in Experiment 1, there was also a significant interaction between the two factors, $F(2, 164) = 6.12$, $MSE = .37$, $p = .003$, $\eta^2 = .07$.

Experiments 3 and 4

The results of Experiments 1 and 2 disagree with those reported in the two previous studies by Dunning and Stern (1992) and Roediger and Payne (1982), who reported no effect of delay on hypermnnesia. Still, they are not in direct conflict with these previous findings. In fact, the present experiments included larger samples of subjects than Dunning and Stern's study did, and they employed longer retention intervals than Roediger and Payne's study did, which may account for the difference in results (see General Discussion). However, there is a possible conflict between the results of present Experiments 1 and 2 and those reported by Wheeler and Roediger (1985), who across three successive tests observed hypermnnesia after a short delay but no hypermnnesia after a prolonged delay.

There are several methodological differences between the present experiments and the one reported in Wheeler and Roediger (1985). For instance, Wheeler and Roediger employed a short delay of 11.5 min and a long delay of 1 week, whereas, in the present experiments, the short delay lasted 3 min and the long delay 24 h; Wheeler and Roediger tested subjects in groups, ranging in size from 3

to 9, whereas we tested subjects individually; and Wheeler and Roediger presented 60 items for study, which were shown in the same serial order to all subjects, whereas we presented 44 items in a random order. We speculate that these differences are not at the core of the conflict in results.

A more critical methodological difference between studies may be recall format. Whereas in the present study, free recall tests were applied across the series of recall tests, Wheeler and Roediger employed forced recall tests. In these tests, subjects were given recall sheets with a separate line for each single to-be-recalled item and were asked to recall as many items as possible. In particular, if unable to remember all studied items, subjects should fill in the remaining spaces with their best guesses. Although there is evidence that recall format does not influence hypermnesia after a short delay (Roediger & Payne, 1985), an influence after long delay can not be excluded. For instance, allowing subjects to fill in the remaining spaces of a recall sheet with their best guesses may not much reduce subjects' effort to recall further previously studied items after a short delay, when recall is still relatively easy. But it may do so after a prolonged delay when recall becomes more demanding. If so, free and forced recall may lead to similar hypermnesia after short delay, but free recall may lead to higher hypermnesia than forced recall after prolonged delay. Experiments 3 and 4 examined the possible role of recall format for hypermnesia directly.

There were two goals with Experiments 3 and 4. The goal of Experiment 3 was to replicate Wheeler and Roediger's (1992) finding of decreased hypermnesia with delay using forced recall at test, the same number of study items, and the same delay intervals as were used in the previous study. The goal of Experiment 4 then was to examine whether forced recall was critical for the results of Experiment 3 and whether results would change if a free recall format was applied at test. If recall format was the critical difference between the present Experiments 1 and 2 and the experiment reported in Wheeler and Roediger (1985), then the results of Experiment 3 using forced recall should replicate those of Wheeler and Roediger (1985) and the results of Experiment 4 using free recall should replicate those of Experiments 1 and 2.

Experiment 3

Experiment 3 examined the role of delay for hypermnesia, closely following the methods employed by Wheeler and Roediger (1985). Subjects were presented 60 pictures and, after a short delay of 11.5 min or a long delay of 1 week, were asked to recall the study items. In both delay conditions, three successive recall tests were conducted, each test using a forced recall format, thus deviating from the recall format used in Experiments 1 and 2 above. We

expected to replicate the results by Wheeler and Roediger (1985) and find hypermnesia after the short delay but no hypermnesia after the long delay.

Method

Participants On the basis of the analysis of test power in Experiment 1 and because of counterbalancing purposes, 48 students of Regensburg University participated in the experiment ($M = 20.83$ years, range, 19–30 years, 77.1% female). All participants spoke German as native language and took part on a voluntary basis. Again, they received monetary reward or course credit for their participation.

Materials We extended the two study lists (A, B) of Experiments 1 and 2 by adding 12 further line-drawing pictures from the Snodgrass and Vanderwart (1980) norms to each single list (see Appendix B). Doing so, list length became equal to that applied in Wheeler and Roediger (1985). As in this previous study, we did not control for primacy and recency effects in this experiment.

Design and Procedure The design of the experiment was identical to Experiments 1 and 2. Each participant completed two experimental blocks, one in the short and one in the long delay condition in counterbalanced order. Again, the blocks were separated by a 5 min break, in which subjects played Tetris. All 60 line-drawings were presented individually on a screen for 7 s each in random order. With presentation, the label of the drawing was enunciated by the experimenter. Each list was presented once. Following Wheeler and Roediger's (1985) procedure, after study, subjects in both delay conditions recalled as many U.S. presidents (one experimental block) or capital cities (other experimental block) as they could. They were then given a questionnaire on which they guessed how many pictures they had seen, how long each picture had appeared, and the total length of the entire presentation. In addition, they were asked to recall the instructions they had received before item presentation. Doing so, a delay of 11.5 min arose before subjects in the short delay condition were tested. In the long delay condition, subjects were disbanded at this point and asked to return at the same time 7 days later.

At test, subjects completed three successive forced recall tests, each lasting for 7 min, with a 1 min break between tests. The experimenter distributed test sheets, with lines numbered 1 to 60 with the instruction to the subjects to recall as many of the previously studied items as possible, independent of what they may have remembered in possible preceding tests. If they felt unable to remember all 60 objects, they should fill the remaining spaces with their best guesses. If the 60 spaces were not complete after 7 min, the subjects were instructed to fill in the remaining spaces

Table 5 Net recall, item gains, item losses, and intrusions in Experiment 3, separately for the short delay (11.5 min) and the long delay (1 week) delay

Pictures	11.5-min delay		1-week delay	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Forced recall				
Net recall				
Test 1	30.00	7.85	15.29	6.29
Test 2	31.00	8.18	15.79	6.44
Test 3	32.23	8.41	15.96	6.40
Gains and losses				
Gains 1 - 2	2.88	2.18	2.54	1.88
Gains 2 - 3	2.50	1.91	2.04	1.43
Losses 1 - 2	1.88	1.66	2.04	1.88
Losses 2 - 3	1.38	1.38	1.85	1.60
Intrusions				
Test 1	21.54	12.64	33.29	16.07
Test 2	21.88	12.72	36.94	15.37
Test 3	21.10	11.86	36.94	15.36

as quickly as possible, thus again following Wheeler and Roediger's (1985) procedure.

Results

Table 5 shows, separately for the short delay (11.5 min) and the long delay (1 week) conditions, (i) net recall on each single test, (ii) item gains and item losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusions on each single test.

Net recall Net recall data were analyzed by means of a 2×3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). There was a main effect of DELAY, $F(1, 47) = 205.20$, $MSE = 83.17$, $p < .001$, $\eta^2 = .81$, showing typical time-dependent forgetting, and a main effect of TEST, $F(2, 94) = 19.18$, $MSE = 2.62$, $p < .001$, $\eta^2 = .29$, indicating increased recall across tests, i.e., hypermnnesia. In addition, there was a significant interaction between the two factors, $F(2, 94) = 4.14$, $MSE = 3.69$, $p = .019$, $\eta^2 = .08$, suggesting that the test-induced increase in recall varied with delay. In fact, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed a significant main effect in the short delay condition, $F(2, 94) = 18.01$, $MSE = 3.32$, $p < .001$, $\eta^2 = .28$, but no such effect in the long delay condition, $F(2, 94) = 1.93$, $MSE = 2.99$, $p = .171$, $\eta^2 = .04$. In the short delay condition, recall on the second test exceeded that on the first test, $t(47) = 2.75$, $p =$

.008, $d = .12$, and recall on the third test exceeded that on the second test, $t(47) = 3.99$, $p = .001$, $d = .15$.

Item gains and item losses Regarding item gains, a 2×2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed a main effect of TEST, $F(1, 47) = 4.39$, $MSE = 2.09$, $p = .042$, $\eta^2 = .09$, indicating more gains between test 1 and test 2 than between test 2 and test 3. There was no main effect of DELAY, $F(1, 47) = 2.46$, $MSE = 3.06$, $p = .124$, $\eta^2 = .05$ and no interaction between the two factors, $F(1, 47) < 1$. The same analysis for item losses showed no main effect of TEST, $F(1, 47) = 3.78$, $MSE = 1.50$, $p = .058$, $\eta^2 = .07$, no main effect of DELAY, $F(1, 47) = 2.19$, $MSE = 2.28$, $p = .145$, $\eta^2 = .05$, and no interaction between the factors, $F(1, 47) < 1$.

Intrusions As expected from the nature of the forced recall test, intrusion rates were high in this experiment. Intrusions were analyzed by means of a 2×3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). The analysis revealed significant main effects of DELAY, $F(1, 47) = 80.03$, $MSE = 130.00$, $p < .001$, $\eta^2 = .63$, and TEST, $F(2, 94) = 4.73$, $MSE = 23.20$, $p = .011$, $\eta^2 = .09$, showing that, unsurprisingly, there were more intrusions after 1 week than after 11.5 min, and that intrusions differed across tests. There was also a significant interaction between the two factors, $F(2, 94) = 7.00$, $MSE = 16.80$, $p = .001$, $\eta^2 = .13$, reflecting the fact that intrusions in the long, but not the short delay condition, increased across tests.⁵

Discussion

Using forced recall at test, the same number of study items, and the same delay conditions as employed in Wheeler and Roediger (1985), the results of this experiment replicate Wheeler and Roediger's prior finding. While net recall increased significantly across tests in the short delay condition, repeated testing left net recall largely unaffected in the long delay condition. Analysis of item gains and item losses did not reveal significant effects of delay, but there were numerical trends for higher item gains and lower item losses after the short delay, which together created the significant effect of delay on hypermnnesia. Wheeler and Roediger did not report item gains and item losses, so there is no way to compare the present results on gains and losses with the prior work.

⁵It was not possible to push each subject to fill in all missing spaces of the test sheets. As a result, in none of the single conditions, do mean net recall and mean intrusions sum up to 60 (see Table 5).

Experiments 1 and 2 on the one hand and Experiment 3 on the other differ in more than one methodological detail. However, if recall format was the main methodological difference, then the difference in results between Experiments 1, 2 and 3 suggests that recall format can influence the effect of delay on hypermnnesia. Whereas both recall formats may create hypermnnesia after short delay, after long delay, free recall may increase hypermnnesia even further, while forced recall may decrease, or even eliminate, the effect. Experiment 4 examines this proposal directly.

Experiment 4

Experiment 4 repeated Experiment 3 but replaced the forced recall format of Experiment 3 by the free recall format used in Experiments 1 and 2. Thus, again subjects were presented 60 pictures and their labels and, after a short delay of 11.5 min or a long delay of 1 week, were asked to recall the labels of the studied pictures. After the delay, three successive free recall tests were conducted. We expected to replicate the finding of Experiment 3 of significant hypermnnesia after the short delay. However, in contrast to Experiment 3, we expected to find an increase in hypermnnesia in the long delay condition, mainly driven by reduced item losses. Such pattern of results would mimic the findings of Experiments 1 and 2, indicating that, with free recall, delay can increase hypermnnesia. In addition, the same pattern would suggest that recall format can be critical for hypermnnesia and influence whether delay has a beneficial or a detrimental effect on hypermnnesia.

Method

Participants Another 48 students of Regensburg University participated in this experiment ($M = 20.48$ years, range, 18–26 years, 68.8% female). All participants spoke German as native language and took part on a voluntary basis. They received monetary reward or course credit for their participation.

Materials, design, and procedure Materials and design were identical to Experiment 3. The procedure was also largely identical. However, unlike in Experiment 3, a free recall format was employed at test. At the beginning of each test, a blank sheet was distributed with the instruction to report as many of the previously studied items as possible, independent of what they may have remembered in possible preceding tests. Guessing was not encouraged.

Results

Table 6 shows, separately for the short delay (11.5 min) and the long delay (1 week) conditions, (i) net recall on each

Table 6 Net recall, item gains, item losses, and intrusions in Experiment 4, separately for the short delay (11.5 min) and the long delay (1 week) delay

Pictures	11.5-min delay		1-week delay	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Free recall				
Net recall				
Test 1	27.27	8.62	12.38	7.14
Test 2	28.13	9.12	13.31	7.43
Test 3	28.75	9.33	15.00	8.25
Gains and losses				
Gains 1 - 2	2.46	2.21	1.44	1.50
Gains 2 - 3	1.10	1.82	1.92	2.20
Losses 1 - 2	1.60	1.51	.50	.74
Losses 2 - 3	1.44	1.57	.25	.53
Intrusions				
Test 1	1.00	1.19	4.23	4.46
Test 2	1.10	1.32	5.29	4.66
Test 3	1.83	2.66	5.94	5.59

single test, (ii) item gains and item losses between test 1 and test 2, and between test 2 and test 3, and (iii) intrusions on each single test.

Net recall The net recall data were analyzed by means of a 2×3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3). There was a main effect of DELAY, $F(1, 47) = 108.39$, $MSE = 139.40$, $p < .001$, $\eta^2 = .70$, showing typical time-dependent forgetting, and a main effect of TEST, $F(2, 94) = 35.06$, $MSE = 2.90$, $p < .001$, $\eta^2 = .43$, indicating increased recall across tests. In addition, there was a significant interaction between the two factors, $F(2, 94) = 4.21$, $MSE = 2.33$, $p = .018$, $\eta^2 = .08$, indicating that the test-induced increase in recall was larger in the long than the short delay condition. This holds while there was significant hypermnnesia in both delay conditions. Indeed, two follow-up unifactorial ANOVAs with the within-subjects factor of TEST showed a significant main effect of TEST in both the short delay condition, $F(2, 94) = 9.39$, $MSE = 2.82$, $p = .001$, $\eta^2 = 0.17$, and the long delay condition, $F(2, 94) = 35.31$, $MSE = 2.41$, $p < .001$, $\eta^2 = .43$. In the short delay condition, recall on the second test exceeded that on the first test, $t(47) = 2.70$, $p = .010$, $d = .10$, but recall on the third test did not exceed that on the second, $t(47) = 1.96$, $p = .056$, $d = .07$. In contrast, in the long delay condition, both recall on the second test exceeded that on the first test, $t(47) = 4.05$, $p < .001$, $d = .13$, and recall on the third test exceeded that on the second, $t(47) = 5.43$, $p < .001$, $d = .22$.

Item gains and item losses Regarding item gains, a 2×2 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) revealed a main effect of DELAY, $F(1, 47) = 6.36$, $MSE = 2.76$, $p = .015$, $\eta^2 = .12$, indicating that there were more gains after the short than the long delay. The main effect of TEST and the interaction between the factors were not significant, both $F's(1, 47) < 2.85$, $MSE's > 2.16$, $p's > .056$, $\eta's^2 < .08$. The same ANOVA for item losses showed a significant main effect of DELAY, $F(1, 47) = 51.09$, $MSE = 1.23$, $p < .001$, $\eta^2 = .52$, suggesting that item losses in the short delay condition exceeded item losses in the long delay condition. The main effect of TEST and the interaction between the two factors were nonsignificant, both $F's(1, 47) < 1.63$, $MSE's > 1.28$, $p's > .207$, $\eta's^2 < .03$.

Intrusions A 2×3 ANOVA with the within-subjects factors of DELAY (short, long) and TEST (test 1, test 2, test 3) revealed significant main effects of DELAY, $F(1, 47) = 41.68$, $MSE = 25.47$, $p < .001$, $\eta^2 = 0.47$, and TEST, $F(2, 94) = 14.16$, $MSE = 2.74$, $p < .001$, $\eta^2 = .23$, showing that there were more intrusions after 1 week than after 11.5 min, and that intrusions increased across tests. The interaction was not significant, $F(2, 94) = 2.17$, $MSE = 3.12$, $p = .120$, $\eta^2 = .04$.

Additional analyses

The results of Experiments 3 and 4 above suggest similar hypermnesia for free and forced recall after the short retention interval, but different hypermnesia for the two recall formats after the long retention interval. Statistical analyses support this suggestion.

Short delay conditions Regarding net recall, a 2×3 ANOVA with the between-subjects factor of RECALL FORMAT (forced, free) and the within-subjects factor of TEST (test 1, test 2, test 3) revealed a main effect of TEST, $F(2, 188) = 26.87$, $MSE = 3.07$, $p < .001$, $\eta^2 = .22$, indicating hypermnesia. The effect of RECALL FORMAT, $F(1, 94) = 3.06$, $MSE = 215.80$, $p = .084$, $\eta^2 = .03$, and the interaction, $F(2, 188) = 1.24$, $MSE = 3.07$, $p = .291$, $\eta^2 = .01$, were nonsignificant. Regarding both item gains and item losses, a 2×2 ANOVA with the factors of RECALL FORMAT (forced, free) and TEST (test 2, test 3) showed no main effects and no interactions, all $F's(1, 94) < 2.98$, $MSE's > 1.79$, $p's > .088$, $\eta's^2 < .03$.

Long delay conditions Regarding net recall, a 2×3 ANOVA with the factors of RECALL FORMAT and TEST revealed a main effect of TEST, $F(2, 188) = 24.22$,

$MSE = 2.70$, $p < .001$, $\eta^2 = .21$, but no main effect of RECALL FORMAT, $F(1, 94) = 2.26$, $MSE = 142.72$, $p = .136$, $\eta^2 = .02$. The interaction was significant, $F(2, 188) = 29.40$, $MSE = 2.70$, $p < .001$, $\eta^2 = .09$, pointing to higher hypermnesia with free recall testing. Regarding item gains, a 2×2 ANOVA with the factors of RECALL FORMAT and TEST revealed a main effect of RECALL FORMAT, $F(1, 94) = 4.58$, $MSE = 3.96$, $p = .035$, $\eta^2 = .05$, with more item gains with forced than free recall testing. There was no main effect of TEST, $F(1, 94) < 1$, but a significant interaction, $F(1, 94) = 4.85$, $MSE = 2.37$, $p = .030$, $\eta^2 = .05$. Regarding item losses, the same ANOVA showed a main effect of RECALL FORMAT, $F(1, 94) = 57.62$, $MSE = 2.06$, $p = .001$, $\eta^2 = .38$, with less item losses with free than forced recall testing. There was no main effect of TEST and no interaction, all $F's(1, 94) < 1.65$, $MSE's > 1.40$, $p's > .203$, $\eta's^2 < .02$.

Discussion

The results of the experiment demonstrate hypermnesia in both delay conditions, but the effect was larger in the long than the short delay condition. This effect of delay was driven by a reduction in item losses across tests in the long delay condition. The reduction in item losses was numerically larger than the simultaneously observed reduction in item gains, which is why an increase in net recall arose with delay. Intrusions also increased with delay. Again, this increase could reflect a more liberal recall threshold in the long than the short delay condition and thus, in principle, could underlie the observed increase in hypermnesia. However, as already mentioned in the discussion of Experiment 1, there is reason to reject such proposal, because loosening the criterion across tests should increase item gains more than affecting item losses, which is not what the present results show. All in all, the results of Experiment 4 thus mimic those of Experiments 1 and 2 above and indicate that, with free recall at test, delay can increase hypermnesia.

The results of Experiment 4 clearly differ from those of Experiment 3. While the results of Experiment 4 show that usage of a free recall format can lead to an increase in hypermnesia with delay, the results of Experiment 3 show that usage of a forced recall format can lead to a decrease with delay. This holds while the two recall formats lead to similar results after short retention interval. Prior work already demonstrated that recall format has no major influence on hypermnesia after short delay (Roediger & Payne, 1985). The present results support this equivalence proposal, but they also show that the proposal does no longer hold when retention interval is prolonged.

General discussion

The results of the present experiments replicate prior work by showing that net recall increases with multiple tests, and that this effect can be larger with pictures as stimuli than with words (e.g., Erdelyi and Becker, 1974; Madigan & Lawrence, 1976). Going beyond the prior work, the present results show that the delay between study and test can influence hypermnnesia. Indeed, when free recall was used as testing format, hypermnnesia was larger after a long delay of 24 h (Experiments 1 and 2) or 7 days (Experiment 4) than after a short delay of 3 min (Experiments 1 and 2) or 11.5 min (Experiment 4). Moreover, in all three experiments, the delay-induced influence on hypermnnesia was driven mainly by differences in item losses, with less previously recalled items being forgotten between tests in the long delay than the short delay condition. There was no increase in item gains with delay. Together, these results indicate that a longer delay between study and test can increase hypermnnesia and does so primarily by reducing the forgetting across recall tests.

The present experiments also show that recall format can influence the effect of delay on hypermnnesia. Employing forced recall (Experiment 3) instead of free recall (Experiment 4) at test, the results first of all showed equivalent hypermnnesia in the two recall formats after short delay, which replicates prior work by Roediger and Payne (1985). Increasing the delay, however, led to nonequivalent hypermnnesia effects, with an increase in hypermnnesia with free recall testing and a decrease with forced recall testing (see also Wheeler & Roediger, 1985). The decrease was reflected in both reduced item gains and increased item losses, although both effects were present numerically only but not statistically. These findings suggest a role of recall format in hypermnnesia, indicating that different mechanisms may mediate the effects of repeated testing in the two recall conditions.

Implications of the free recall results for accounts of hypermnnesia

The present free recall results on net recall are consistent with the changes in cue set hypothesis (Raaijmakers & Shiffrin, 1980; Roediger & Thorpe, 1978), the retrieval strategy hypothesis (Erdelyi & Becker, 1974; Mulligan, 2001), and the retrieval practice hypothesis (Hogan & Kintsch, 1971; Roediger & Karpicke, 2006). On the basis of the changes in cue set hypothesis, hypermnnesia is expected to increase with delay if, after delay-induced contextual drift, retrieval of some items reactivates the study context and such reactivation is not completed at the end of the first test but extends to later recall tests. Because context reactivation can change the cue set that people use to sample

and recover items, in such case it can induce alternative retrieval routes on later recall tests, which may enhance item gains and increase hypermnnesia. According to the retrieval strategy hypothesis, enhanced organization across repeated tests leads to hypermnnesia, improving recall by increasing item gains and reducing item losses. If such organization was further advanced after longer delay, for instance, because retrieval after delay becomes more challenging, then item gains should be further enhanced with delay and item losses be limited, again increasing hypermnnesia. Also the retrieval practice hypothesis can account for the present free recall results. Because retrieval practice should be more demanding after longer than after shorter delay, and retrieval practice effects have been shown to be particularly strong if retrieval practice is demanding, retrieval after longer delay may lead to enhanced hypermnnesia by reducing the forgetting of the initially recalled items. In contrast, the present finding of increased hypermnnesia with delay is not easily reconciled with the cumulative recall hypothesis (Roediger & Challis, 1989; Roediger et al., 1982). This hypothesis claims that study conditions producing high levels of asymptotic recall should induce more hypermnnesia than conditions producing lower levels of recall, which is the opposite of what the present results show.

Although the changes in cue set hypothesis, the retrieval strategy hypothesis, and the retrieval practice hypothesis are consistent with the finding of increased hypermnnesia with delay, they differ in the degree to which they can explain the observed presence of a delay effect in item losses and the observed absence of a delay effect in item gains. The changes in cue set hypothesis is largely focused on item gains and thus explains the effect of delay mainly by attributing it to enhanced item gains. The retrieval strategy hypothesis makes assumptions about both item gains and item losses and suggests a beneficial effect of delay on item gains and a detrimental one on item losses. Finally, the retrieval practice hypothesis focuses mainly on item losses and thus explains the effect of delay primarily by a reduction in item losses. The present finding that, with free recall as testing format, delay increases hypermnnesia mainly by reducing item losses thus favors the retrieval practice hypothesis, indicating that retrieval practice effects can contribute to hypermnnesia and do so particularly when the delay between study and test is increased.

The finding of Experiments 1, 2 and 4 that the increase in hypermnnesia with delay is due to a reduction in item losses, arose by analyzing absolute differences in recall levels between tests, which is typical for prior work on hypermnnesia (e.g., Dunning and Stern, 1992; Mulligan, 2005; Wheeler and Roediger, 1985; but see Goernert, Widner, & Otani, 2007). However, one may also take a different view on the issue. Indeed, because after prolonged delay, fewer items are recalled than after short delay (see Tables 2, 3, 4, 5

and 6), one could argue that there are also fewer items to drop between tests after the longer delay, which raises the question of whether the results reported in Experiments 1, 2 and 4 above would replicate if a measure of proportion of items dropped was employed for analysis. Using such proportion measure, corresponding analyses showed that the pattern of results reported above indeed replicates and item losses remain reduced with delay in each of the three experiments.⁶ The main results of the present study thus do not depend on whether absolute or proportion measures are used for analysis.

Relation of the free recall results to prior work

The present finding that, with free recall as testing format, longer delay increases hypermnnesia disagrees with the results of two previous studies that also examined the role of delay between study and test for hypermnnesia and found no effect of delay. In the one study, Dunning and Stern (1992; Experiment 2) investigated hypermnnesia in eyewitness memory using films about crime scenes as stimulus material and employing a single experiment with, on average, less than ten subjects per condition. The present study reports the results of three experiments with at least 42 subjects per condition, using both words and pictures as stimulus material. While it cannot be excluded that stimulus material can affect hypermnnesia results (e.g., Ballard, 1913), it appears more likely that the difference in results between the previous and present studies has to do with the difference in statistical power, in particular, as the statistical power employed in Dunning and Stern's experiment should have been too low to detect a possible effect of delay on hypermnnesia (see also Methods of Experiment 1 above). In the other study, Roediger and Payne (1982) reported another single experiment, in which delay was manipulated by conducting the initial recall test immediately after study, or after subjects read a prose passage for 18 min. Because, in contrast to the present study, delay did not induce any time-dependent forgetting in this previous study, the difference in results between the previous and the present studies may reflect the difference in degree to which the employed delay manipulations were effective. The present results thus are not in direct conflict with the results from these two previous studies and may rather indicate that, in order to observe an effect of delay on hypermnnesia, sufficient statistical power is required and a delay interval that induces robust time-dependent forgetting.

⁶Regarding item losses, 2×2 ANOVAs with the within-subjects factors of DELAY (short, long) and TEST (test 2, test 3) showed a significant main effect of DELAY in all three experiments, $F's > 6.88$, $MSE's < 40.17$, $p's < .012$, $\eta's^2 > .14$, indicating that losses were indeed reduced after prolonged delay. Regarding item gains, the pattern of results reported above also replicated.

The present results are in line with the testing effect literature, which shows that retrieval practice can improve recall of practiced items and does so even more if retrieval practice is demanding (e.g., Bäuml et al., 2014; Carpenter, 2011; Halamish & Bjork, 2011; Pyc & Rawson, 2009). In particular, this literature has shown that retrieval practice can reduce the forgetting of practiced items and thus enhance long-term retention e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006). Using a different paradigm, the present experiments reveal a similar pattern by showing that, after longer delay between study and test, retrieval practice on an initial recall test can reduce the forgetting of practiced items on subsequent recall tests relative to a short delay condition. Enhanced hypermnnesia after longer delay can thus serve as another demonstration of the role of difficulty of retrieval practice task for beneficial effects of retrieval practice. Moreover, on the basis of the testing effect literature, the present findings also suggest that hypermnnesia may be enhanced whenever the initial test is demanding. Hypermnnesia may thus be increased not only after longer delay, but also after a change in context between study and test, or in the presence of interference. Future work may investigate the issue in more depth.

Results from several recent studies suggest that, after longer delay between study and test, selective retrieval can improve recall of other items (e.g., Bäuml & Schlichting, 2014; Wallner & Bäuml, 2017). The finding was interpreted as evidence that, after a delay and induced context change (e.g., Bower, 1972; Estes, 1955), retrieval of some first items reactivates the items' study context, which then serves as a retrieval cue for the remaining items and improves recall performance (see also Bäuml & Samenieh, 2012; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988). Critically, if context reactivation was still incomplete at the end of the first recall test, reactivation might still operate on the subsequent test, leading to retrieval of further items and increased hypermnnesia. The present findings do not show such an increase in item gains, however. Increases in recall due to reactivated context thus may be largely restricted to the first test and not easily extend to subsequent recall tests.

The results of Experiments 1, 2 and 4, which address the role of delay between study and test for hypermnnesia, complement prior work by Mulligan (2006) who, also employing free recall tests, investigated the role of inter-test delay for hypermnnesia. Using free recall, many previous studies already showed hypermnnesia after very long inter-test delays, like days, weeks, months, or even a year (e.g., Campbell, Nadel, Duke, & Ryan, 2011; Erdelyi, 1996), but Mulligan was the first to explicitly compare hypermnnesia under different inter-test delay conditions. In this study, hypermnnesia was found to increase with inter-test delay (7 min vs. immediate recall), with the increase being due to increases in item gains and hardly to reductions in item losses.

Together with the present results, these findings suggest that, with free recall testing, both delay between study and test and inter-test delay can influence hypermnnesia, though in different ways. Whereas increased delay between study and test seems to affect mainly item losses (present study), increased inter-test delay seems to affect mainly item gains (Mulligan, 2006). Future work may investigate the possible interaction between the two delay factors and examine whether the present free recall results generalize to conditions in which delay between tests is increased, and the results by Mulligan generalize to conditions in which delay between study and test is increased.

Free recall versus forced recall testing

Erdelyi and Becker (1974) introduced forced recall testing in the hypermnnesia literature in order to control for possible criterion changes across successive recall tests. Comparing the effects of forced recall and free recall on hypermnnesia, the results of several studies, however, reported equivalent hypermnnesia effects, first of all indicating that changes in response criteria may play little, if any, role in hypermnnesia (e.g., Roediger & Payne, 1985). While this prior work focused on short delay conditions between study and test, the present study includes both short and long delay conditions. Doing so, the results of the present Experiments 3 and 4 show nonequivalent hypermnnesia effects for the two recall formats after prolonged delay, suggesting an effect of recall format on response criterion. However, it is unlikely that differences in response criteria mediated the difference in hypermnnesia results. Indeed, if the nonequivalence between free and forced recall was caused by loosened criterion with free recall testing, then the increase in hypermnnesia with delay observed with free recall testing should have been accompanied by an increase in item gains rather than by a reduction in item losses, which is not what the present results show.

The results rather suggest that different mechanisms may mediate hypermnnesia after delay with forced versus free recall testing. Whereas the increase in hypermnnesia with free recall testing is in line with the retrieval practice hypothesis but is inconsistent with the cumulative recall hypothesis (see above), the opposite is true for the decrease in hypermnnesia with forced recall testing. Indeed, because (cumulative) recall levels after longer delay are lower than after short delay (see Table 5), the finding of decreased hypermnnesia after delay agrees with the cumulative recall hypothesis, which assumes that study conditions producing low levels of asymptotic recall should induce less hypermnnesia than conditions producing higher levels of recall. The observed numerical (though not statistical) reduction in item gains with delay fits also with this view. Answering the questions of why different

mechanisms may mediate hypermnnesia with forced recall than with free recall testing, and why repeated testing after delay reduces item losses with free recall only, is beyond the scope of the present study and future work is required to address these issues. Such work may improve not only our understanding of hypermnnesia but also of the relation between free and forced recall testing in general.

Conclusions

In four experiments, we showed that hypermnnesia varies with the delay between study and test. When free recall was used at test, hypermnnesia increased with delay and the effect was driven mainly by reduced item losses between tests. This result fits with the view that retrieval practice reduces the forgetting of practiced items and does so even more if retrieval is demanding, i.e., after the long delay, which suggests a link between hypermnnesia and the testing effect. When forced recall was used at test, hypermnnesia decreased with delay and was even absent after longer delay. These findings indicate that recall format can influence hypermnnesia and different mechanisms may mediate the effects of repeated testing with free and forced recall testing.

Acknowledgments This research is part of L.A. Wallner's dissertation.

Appendix A

Table 7 Words (and also line-drawing pictures) presented to participants in Experiments 1–4

Set 1					
parrot	bat	pear	lamp	radio	mirror
whale	bird	plank	mixer	girl	carpet
slug	bug	letter	strawberry	rain	bowl
baby	bride	violin	rainbow	slide	seesaw
camel	dentist	cheese	nail	shovel	tent
zebra	boy	bone	pizza	ship	sheet
bread	elephant	helmet	cup	bag	queen
pirate	broom	tie	puzzle	pants	moose
Set 2					
cowboy	panda	brush	microscope	plate	bomb
crab	octopus	egg	nest	stairs	bridge
fish	penguin	feather	patch	clock	chain
ghost	policeman	glass	rocket	colcannon	cross
king	chest	towel	rose	balloon	cactus
lizard	desert	comb	box	bathtub	butter
llama	lion	tape	rope	belt	doctor
monkey	drill	ladder	scarf	bench	fireman

Appendix B

Table 8 Labels of line-drawing pictures additionally presented to participants in Experiments 3 and 4

Set 1					
plane	book	apple	bus	flower	dog
ant	axe	cake	mushroom	ear	moon
Set 2					
carrot	pipe	ruler	ring	banana	scissors
key	refrigerator	toaster	flag	chair	cat

References

- Bäumli, 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*, 1568–1581.
- Bäumli, K.-H. T., & Samenih, A. (2012). Selective memory retrieval can impair and improve retrieval of other memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 388–394.
- Bäumli, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition*, *132*, 16–21.
- Ballard, P. B. (1913). Obliviscence and reminiscence. *British Journal of Psychology Monograph Supplements*, *1*, 1–82.
- Belmore, S. M. (1981). Imagery and semantic elaboration in hypermnnesia for words. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 191–203.
- Bergstein, J., & Erdelyi, M. (2008). Recognition hypermnnesia: How to get it. *Memory*, *16*, 689–702.
- Bluck, S., Levine, L. J., & Lulhere, T. M. (1999). Autobiographical remembering and hypermnnesia: A comparison of older and younger adults. *Psychology and Aging*, *14*, 671–682.
- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. *Coding Processes in Human Memory*, 85–123.
- Campbell, J., Nadel, L., Duke, D., & Ryan, L. (2011). Remembering all that and then some: Recollection of autobiographical memories after a 1-year delay. *Memory*, *19*, 406–415.
- 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*, 1547–1552.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*, 2nd ed. Hillsdale: Erlbaum.
- Dunning, D., & Stern, L. B. (1992). Examining the generality of eyewitness hypermnnesia: A close look at time delay and question type. *Applied Cognitive Psychology*, *6*, 643–657.
- Erdelyi, M. H. (1996). *The recovery of unconscious memories: Hypermnnesia and reminiscence*. Chicago: The University of Chicago Press.
- Erdelyi, M. H., & Becker, J. (1974). Hypermnnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, *6*, 159–171.
- Erdelyi, M. H., & Kleinbard, J. (1978). Has Ebbinghaus decayed with time? The growth of recall (hypermnnesia) over days. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 275–289.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*, 145–154.
- 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*, 175–191.
- Goernert, P. N., Widner, R. L. Jr., & Otani, H. (2007). Classification accuracy across multiple tests following item method directed forgetting. *Quarterly Journal of Experimental Psychology*, *60*, 1178–1186.
- 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*, 801–812.
- 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*, 562–567.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923–941.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269–299.
- Kelley, M. R., & Nairne, J. S. (2003). Remembering the forgotten? Reminiscence, hypermnnesia and memory for order. *The Quarterly Journal of Experimental Psychology*, *56*, 577–599.
- Madigan, S. (1976). Reminiscence and item recovery in free recall. *Memory and Cognition*, *4*, 233–236.
- Madigan, S., & Lawrence, V. (1980). Factors affecting item recovery and hypermnnesia in free recall. *The American Journal of Psychology*, *93*, 489–504.
- McDaniel, M. A., Moore, B. A., & Whiteman, H. L. (1998). Dynamic changes in hypermnnesia across early and late tests: A relational/item-specific account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 173–185.
- Mensink, G. J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455.
- 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*, 49–62.
- Mulligan, N. W. (2001). Generation and hypermnnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 436–450.
- Mulligan, N. W. (2002). The emergent generation effect and hypermnnesia: Influences of semantic and nonsemantic generation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 541–554.

- Mulligan, N. W. (2005). Total retrieval time and hypermnesia: Investigating the benefits of multiple recall tests. *Psychological Research Psychologische Forschung*, *69*, 272–284.
- Mulligan, N. W. (2006). Hypermnesia and total retrieval time. *Memory*, *14*, 502–518.
- Nelson, T. O., & MacLeod, C. M. (1974). Fluctuations in recall across successive test trials. *Memory and Cognition*, *2*, 687–690.
- Otani, H., & Griffith, J. D. (1998). Hypermnesia for prose. *The Journal of General Psychology*, *125*, 147–155.
- Otani, H., Widner, R. L., Whiteman, H. L., & Louis, J. P. S. (1999). Hypermnesia: The role of multiple retrieval cues. *Memory and Cognition*, *27*, 928–934.
- Payne, D. G. (1986). Hypermnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 16–29.
- Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, *101*, 5–27.
- Payne, D. G., & Roediger, H. L. I. (1987). Hypermnesia occurs in recall but not in recognition. *The American Journal of Psychology*, *100*, 145–165.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*, 437–447.
- Raaijmakers, J. G., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In Ross, B. (Ed.) *The psychology of learning and motivation* (Vol. 14). Academic Press: Elsevier Inc.
- Roediger, H. L., & Challis, B. H. I. (1989). Hypermnesia: Improvements in recall with repeated testing. In Izawa, C. (Ed.) *Current issues in cognitive processes: The tulane flowerree 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*, 249–255.
- Roediger, H. L., & Payne, D. G. I. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 66–72.
- Roediger, H. L., & Payne, D. G. I. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory and Cognition*, *13*, 1–7.
- 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*, 635–655.
- Roediger, H. L., & Thorpe, L. A. I. (1978). The role of recall time in producing hypermnesia. *Memory and Cognition*, *6*, 296–305.
- Shapiro, S. R., & Erdelyi, M. H. (1974). Hypermnesia for pictures but not words. *Journal of Experimental Psychology*, *103*, 1218–1219.
- 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*, 174–214.
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175–184.
- Wallner, L., & Bäuml, K.-H. T. (2017). Beneficial effects of selective item repetition on the recall of other items. *Journal of Memory and Language*, *95*, 159–172.
- 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*, 240–245.
- Wilkinson, A. C., & Koestler, R. (1983). Repeated recall: A new model and tests of its generality from childhood to old age. *Journal of Experimental Psychology: General*, *112*, 423–451.