



Personal reminders: Self-generated reminders boost memory more than normatively related ones

Di Zhang¹ · Jonathan G. Tullis¹

Accepted: 23 November 2020 / Published online: 7 January 2021
© The Psychonomic Society, Inc. 2021

Abstract

People generate reminders in a variety of ways (e.g. putting items in special places or creating to-do lists) to support their memories. Successful reminders can result in retroactive facilitation of earlier information; in contrast, failures to remind can produce interference between memory for related information. Here, we compared the efficacy of different kinds of reminders, including participant's self-generated reminders, reminders created by prior participants, and normatively associated reminders. Self-generated reminders boosted memory for the earlier target words more than normatively associated reminders in recall tests. Reminders generated by others enhanced memory as much as self-generated reminders when we controlled output order during recall. The results suggest that self-generated reminders boost memory for earlier studied information because they distinctly point towards the target information.

Keywords Reminding · Memory · Metacognition · Generation

Reminders help us thrive in a complex and changing environment by connecting relevant past knowledge to current experiences. Meeting new people can remind us of people we already know, current music can remind us of oldies, and novel problems can remind us of ones we have previously solved. Reminders, stimulus-guided retrievals of prior episodes, can bring relevant past experiences to bear to solve novel problems and thereby link information separated in time (Benjamin & Ross, 2011). Here, we examine whether learners can generate their own reminders successfully and explore what makes a reminder effective.

Reminders may underlie a wide range of cognitive functions, from basic memory processes (e.g., Tullis, Benjamin, & Ross, 2014) to complex problem-solving processes (e.g., Reeves & Weisberg, 1994). Reminders were first shown to impact memory more than 40 years ago, when studying novel information retroactively facilitated memory for earlier related information (Bruce & Weaver, 1973). Reminders have since been implicated in many memory processes, including recency judgments (Hintzman, 2010; Tzeng & Cotton, 1980;

Wahlheim & Jacoby, 2013; Winograd & Soloway, 1985), the spacing effect (Benjamin & Tullis, 2010), judgments of the spacing between repeated episodes (Hintzman, Block, & Summers, 1973), frequency judgments (Hintzman, 2004), list discrimination (Jacoby, Wahlheim, & Yonelinas, 2013), memory for source information (Tullis & Benjamin, 2020), and memory for change (Jacoby, Wahlheim, & Kelley, 2015). Reminders have further been theorized to affect higher order cognition, including categorization (Medin & Schaffer, 1978; Ross, Perkins, & Tenpenny, 1990), interpretation of ambiguous events (Ross & Bradshaw, 1994; Tullis, Braverman, Ross, & Benjamin, 2014), generalization (Gick & Holyoak, 1983; Ross & Kennedy, 1990), and transfer (Tullis & Goldstone, 2016).

The mnemonic impact of studying related information depends on whether or not reminders occur. When related information reminds learners of previously studied information, retroactive facilitation boosts recall (Tullis, Benjamin, et al., 2014); when learners fail to recollect the change that happened during study, related information can cause proactive or retroactive interference between related targets at test (Jacoby & Wahlheim, 2013). Therefore, knowing what causes reminders is important to understanding how related information will impact subsequent recall of that information. Two central factors have been theorized to impact the occurrence of reminders. First, the timing between related episodes may determine whether reminders occur. Long lags between

✉ Di Zhang
dizhang@email.arizona.edu

¹ Department of Educational Psychology, University of Arizona, 1430 E. Second St., Tucson, AZ 85721, USA

related episodes may reduce the probability of reminding because learners forget the earlier instance before the later episode or because intervening items interfere with the reminding. Evidence suggests that long lags prevent remindings between related episodes. For example, the mnemonic benefits of presenting related information disappear when related items are spaced far apart (e.g., Benjamin & Tullis, 2010; Hintzman, Summers, & Block, 1975). Even the mnemonic benefits of repetitions decrease when repetitions are spaced at very long lags (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The timing between related episodes, then, is one important factor that can determine the impact of remindings on memory.

Second, remindings are likely driven by the similarity or association between related episodes. For example, in problem solving research, the superficial similarity of the novel problem to prior problems drives the probability of remindings (Ross, 1984, 1987). The similarity of the superficial context between episodes (i.e., an unrelated background picture) also affects how probable remindings are (Tullis, Braverman, et al., 2014). Beyond the timing of episodes and the relationship between episodes, we know little about what makes a reminder effective. Here, we test the effectiveness of different kinds of reminders on memory, including reminders that learners generate for themselves, which allows us to examine characteristics underlying successful reminders.

Self-generated reminders

People generate reminders throughout their daily lives. People create to-do lists to remind them of their plans for the day, generate file names to remind them of the contents of the file, and take photos to remind them of specific experiences (see Tullis & Finley, 2018). In fact, in a survey about the types of reminders people generate, 100% of people reported putting items in a special place to remind them of something, 97% reported writing notes to themselves, 93% reported writing shopping lists, and 53% reported using a timer as reminders (Harris, 1980).

The specific effectiveness of self-generated reminders is largely untested (but see evidence about the ineffectiveness of self-generated reminders to take medication; Osterberg & Blaschke, 2005). While the efficacy of self-generated *reminders* has not been tested, much research on self-generated *test cues* shows that they support recall more effectively than test cues provided by others. For example, when learners generate their own descriptions of target stimuli, these descriptions cue memory very effectively, even for long lists of items and for long retention intervals (Bäckman & Mäntylä, 1988; Hunt & Smith, 1996; Mäntylä, 1986; Mäntylä & Nilsson, 1983, 1988; Tullis & Benjamin, 2015a, 2015b). The effectiveness of self-generated test cues derives, at least partially, from their distinctiveness (Tullis & Fraundorf, 2017): Learners produce cues that point to a small

number of potential targets. Reducing the number of potential targets that a cue leads to can guide access to the appropriate target item, exclude nontargets from retrieval, and ultimately bolster the efficacy of the cue (Hunt, 2006; Waddill & McDaniel, 1998; Watkins & Watkins, 1976). In the current experiments, we will examine whether distinctiveness affects the efficacy of reminders.

The primary distinction between test cue generation and reminding is the intentionality of retrieval. In research examining the effectiveness of test cue generation, learners are given their cues during cued recall tests when they are actively searching memories for studied episodes (i.e., they may be in a distinct “retrieval mode”; Lepage, Ghaffar, Nyberg, & Tulving, 2000). However, reminders can occur during study when participants are trying to encode new information (and not solely trying to retrieve any studied episodes). For example, novel problems can remind learners of previously answered problems during math instruction (e.g., Bassok & Holyoak, 1989; Ross, 1987) and novel exemplars can remind learners of prior studied exemplars during category learning (Ross, 1996; Ross et al., 1990). In these scenarios, reminders bring prior information to mind when learners are just trying to encode new information or solve new problems. Reminders need to *trigger* retrieval of prior encoded items to be effective, but tests cues do not because learners are already in a retrieval mode.

Across three experiments, we tested the effectiveness of different reminders on memory for earlier episodes. Learners studied a list of words one at a time; the list contained the target words followed by their own reminders, normative reminders, or others’ reminders. Finally, learners took a free recall test (Experiment 1), a recognition test (Experiment 2), or an extra-list cued recall test (Experiment 3) to assess their memory for the target items. We examined three specific questions in each of the three experiments. First, we examined whether studying related words boosts memory for earlier normatively related targets to test whether related information reminds learners of prior information (as has been shown previously; Tullis, Benjamin, et al., 2014). Second, we examined whether self-generated reminders boost memory for earlier information more than normatively related reminders do. Finally, we examined whether self-generated reminders boost memory more than reminders generated by others to test whether idiosyncratic connections and self-reference effects underlie the benefits of self-generated reminders. If self-generated reminders derive effectiveness from connections to personal experiences, self-generated reminders would boost memory for earlier targets more than would reminders generated by others.

Experiment 1

In Experiment 1, we compared the mnemonic effects of four types of reminders (self-generated reminders, reminders

generated by others, normatively related reminders, and unrelated reminders) on free recall of earlier target words. If learners produce particularly effective reminders for their own learning that are tied into their own idiosyncratic knowledge and experiences, self-generated reminders should produce the biggest mnemonic benefit to the target information when compared to the other three types of reminders.

Method

Participants Prior research suggests that reminders have a medium to large effect on free recall (Cohen's d from 0.50 to 1.01; Tullis, Benjamin, et al., 2014). We conducted a power analysis to determine how many participants were needed to detect an effect of 0.5 with 0.8 power, an alpha of 0.05, and a correlation among repeated measurements of zero (to be the most conservative). G*Power suggested we needed 45 participants (Faul, Erdfelder, Lang, & Buchner, 2007), so 45 undergraduate students from the University of Arizona fully participated in exchange for partial course credit for introductory educational psychology courses. Four additional undergraduate students were recruited to provide the initial round of self-generated reminders (but were excluded from all analyses), as described below.

Materials Forty word pairs were selected from those used in Tullis, Benjamin, et al. (2014; Experiment 1). The word pairs were collected from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) and each word in a pair was the strongest associate of the other. Associated pairs were highly related, such that the mean associative strength from the cue to the target equaled 0.27 ($SD = 0.07$). The normatively associated pairs included synonyms (e.g., gross, disgusting), antonyms (e.g., weakness, strength), hyponyms (e.g., dance, ballet), and male and female counterparts (e.g., hero, heroine).

Procedure The experiment was programmed in MATLAB using the Psychophysics Toolbox (Fraundorf et al., 2014). Participants completed the experiment on individual desktop computers while up to three other participants used other computers in the same room. The first participant on each computer saw every possible target word presented one at a time in 50-point black Arial font and were instructed to "Type a word that most makes you think of the top, target word." The first four participants generated their own reminder for each target word; they could not receive any prior participants' reminders (because there were no prior participants) and consequently were excused after this initial phase. Their memories were never tested, and their data are included in no analyses.

All following participants completed three phases (preexposure/generation, delay, and study/test phases), which are depicted in Fig. 1. During the preexposure/generation phase, participants both created their own reminders for the

self-generated condition and rated the relationships for all of the words pairs across all conditions. More specifically, a target word (P1) was presented at the top of the screen and the reminder word (P2) was presented below it. Specific instructions to participants said the following:

In this experiment, we will give you 32 word pairs. Your job is to judge how strongly the bottom word points to the top word (or how strongly the bottom word makes you think of the top word).

You will rate these relationships on a scale of 1 to 4. If you believe the bottom word very strongly directs you to the top word, you would rate it as a 4 (*very strongly directed*), if you believe the bottom word does not point to the top word (or make you think of the top word), you would rate that as a 1 (*very weakly directed*).

For some of the words, we will have you generate a word that points to the top word as strongly as it can for you. After you generate a word that makes you think of the top word, you will rate the strength of its relationship, just like all other pairs.

The top word (i.e., the target, or P1) was always the target word that came from the word norms. The bottom word (i.e., the reminder, or P2) depended upon the condition of the word pair. Word pairs were randomly but equally distributed across four conditions, including normatively related, unrelated, self-generated, and other-generated. In the normatively related condition, the reminder came from the South Florida Free Association Norms, as described above. In the unrelated condition, the reminder was taken from a different, unrelated, and unstudied word pair. In the self-generated condition, participants created their own reminder for a given target. The target word was displayed at the top of the screen below a question that asked, "What word makes you most strongly think of the top word?" After typing in their own reminder, participants rated the strength of the association from their reminder to the target as in the other conditions. Finally, in the other-generated condition, the reminder came from the most recent prior participant who generated a reminder for the given target. A list composed of the most recent participant-generated reminders for each target was stored on each computer; when a target was assigned to the other-generated condition, the most recent reminder generated by a prior participant was pulled from this list.

After rating each of the pairs, participants were told they completed that experiment and started a second experiment. In this delay phase, participants completed a math task for 10 minutes. The math task was inserted between the generation and study phases to allow some forgetting of the items from the generation phase. Further, splitting up the preexposure phase from the study phase with an unrelated task should reduce the impact of the preexposure phase on the study/test phase.

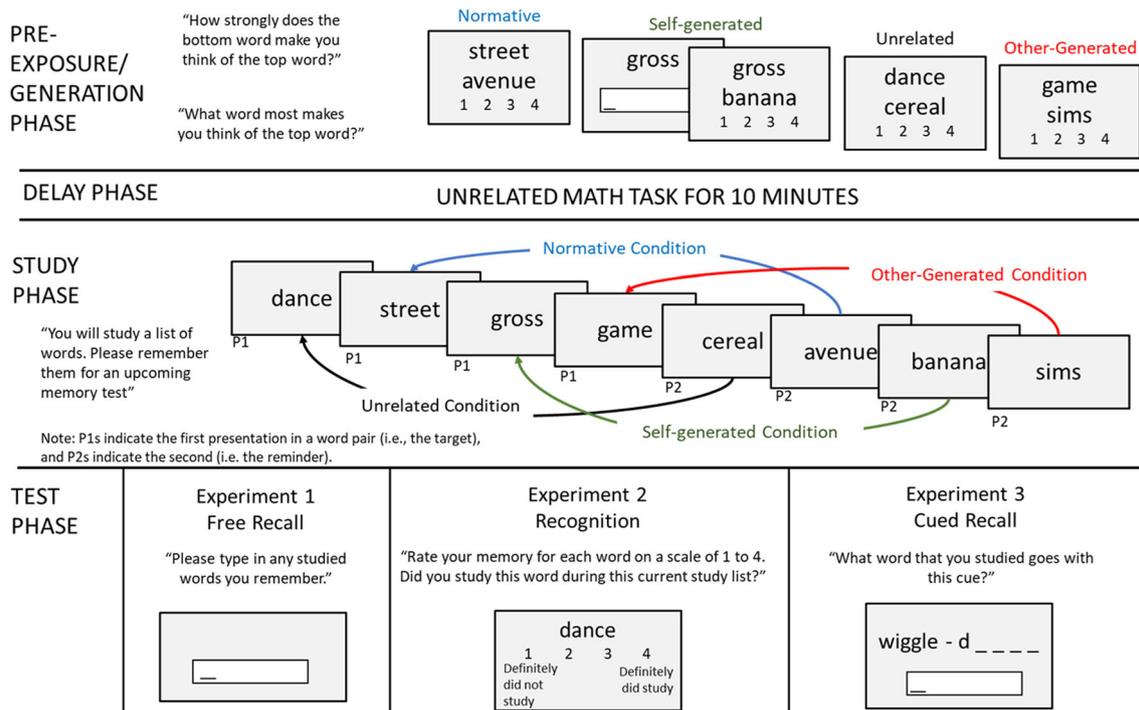


Fig. 1 The procedure and instructions for Experiments 1–3

After participants completed the math task, they were instructed to start a third experiment, which included the study and test phases. Participants were not told about the relationship between the preexposure/generation phase and this study/test phase. The only instructions participants received were, “You will study a list of words. Please remember them for an upcoming memory test.” Participants studied words one at a time in 60-point black Arial font for 5 seconds, followed by 200 milliseconds of blank white screen between presentations. The first word from each pair (P1) and the second word from each word pair (P2) were separated in the study list by three other intervening words from the study list. A lag of three intervening items was chosen to mirror prior research (Tullis, Benjamin, et al. 2014). Lags of three other words allow some forgetting of the first word (P1) before the second word (P2) is shown, but are not too long to prevent reminding. As described above, word pairs were randomly assigned to condition, such that each participant had eight word pairs in each of the self-generated, other-generated, normative, and unrelated conditions for each participant. All words shown during the study list had been presented during the preexposure/generation phase.

List position was controlled across conditions by presenting a word pair from each condition before repeating a condition and randomizing the order within the four conditions. The study list had two primacy buffer words and two recency buffer words that were not related to any study words and were not included in any analyses. After studying a list of 36 items, participants took a free recall test, in which they typed

in any word they could remember from the study list. When participants decided they could remember no more words, they exited the free recall test. They repeated the study/test cycle with a new list of word pairs.

Results

Data from this experiment and those that follow are available on the OSF (https://osf.io/w9m3s/?view_only=3d50437869dd4bf7b7555d5ee797b392).

We first examined the ratings of association that participants made during the preexposure phase, which are shown in Fig. 2. A one-way repeated-measures analysis of variance (ANOVA) showed that participants’ judgments of association differed across the four conditions, $F(3, 132) = 359.81, p < .001, \eta_p^2 = 0.89$. Normatively related word pairs were rated as more strongly associated than unrelated pairs, $t(44) = 31.55, p < .001, d = 4.76$. The judged association between normatively related reminders and self-generated reminders did not significantly differ, $t(44) = 1.41, p = .17, d = 0.21$. Further, the judged association between self-generated reminders and other-generated reminders did not significantly differ, $t(44) = 1.03, p = 0.31, d = 0.15$. The self-generated cues were the same as the normative ones for 12.5% ($SD = 10.87\%$) of the targets. The between-participant correlations among these judgments (and among cued recall) for this experiment and the following experiments are shown in the Appendix.

Next, we examined the proportion of studied words that were correctly recalled, which is shown in Fig. 3. Our

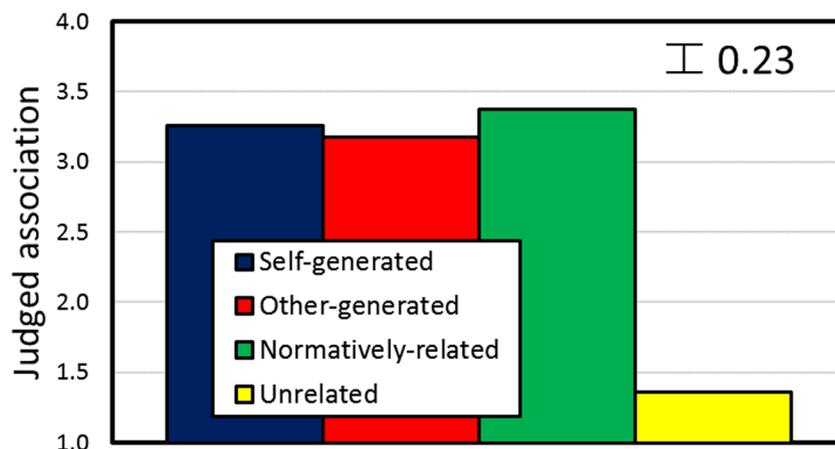


Fig. 2 Judged association of word pairs across the four cue conditions in Experiment 1. The width of the error bars here and on all subsequent graphs indicate the within-subjects 95% confidence interval across conditions (Loftus & Masson, 1994)

hypotheses focus primarily on three specific comparisons. First, we compared recall of the normatively related pairs with the unrelated pairs as a replication of prior research (Tullis, Benjamin, et al., 2014). Second, we compared self-generated cues with normatively related cues to examine whether self-generated reminders produce more robust reminders than normative relationships. Third, we compared recall of self-generated pairs with other-generated pairs to examine whether self-generated reminders specifically benefit oneself (or whether their benefits extend to others' memory).

P1 indicates the words from each word pair that were presented first (the targets), while *P2* indicates the words from each word pair presented second (the reminders). We analyzed *P1*s separately from *P2*s because *P1*s came from the standardized word list, while *P2*s in some conditions were generated by participants. *P1*s, then, should be inherently equally memorable, and any differences found between their recall should be a function of their condition (i.e., whether they were successfully reminded). Recall of *P2*s may be driven by differences in the characteristics of words generated (and may differ due to generation itself; Slamecka & Graf, 1978). A repeated-measures ANOVA on *P1*s revealed a significant effect of condition on recall, $F(3, 132) = 16.01, p < .001, \eta_p^2 = 0.27$. Further, we examined planned contrasts between specific conditions. *P1*s followed by a normatively related word were remembered better than *P1*s followed by

an unrelated word, $t(44) = 3.36, p = .002, d = 0.51$. *P1*s followed by self-generated words were remembered better than *P1*s followed by a normatively related word, $t(44) = 3.47, p = .001, d = 0.52$. Finally, *P1*s followed by a self-generated word were remembered better than *P1*s followed by a word generated by someone else, $t(44) = 3.35, p = .002, d = 0.51$.

While not related to our central research questions, we analyzed the proportion of *P2*s recalled by condition for completeness. A repeated-measures ANOVA revealed a significant effect of condition on recall of *P2*, $F(3, 132) = 25.32, p < .001, \eta_p^2 = 0.37$. *P2*s preceded by normatively related words were remembered better than *P2*s preceded by unrelated words, $t(44) = 7.97, p < .001, d = 1.20$. Recall for self-generated *P2*s and for *P2*s preceded by a normatively related word were not significantly different, $t(44) = 1.53, p = .13, d = 0.23$. Finally, self-generated *P2*s were remembered better than other-generated *P2*s, $t(44) = 4.51, p < .001, d = 0.68$.

Discussion

Experiment 1 revealed a strong effect of the kind of reminder (*P2*) on memory for earlier related items (*P1*s). Normatively related reminders boosted recall of *P1* compared with the unrelated condition, which replicates prior research (Tullis, Benjamin, et al., 2014). Further, self-generated reminders benefited memory

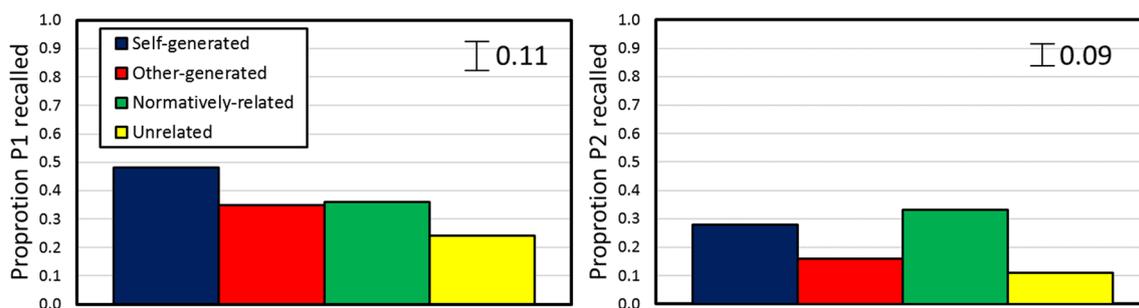


Fig. 3 Proportion of *P1*s (left panel) and *P2*s (right panel) recalled in Experiment 1

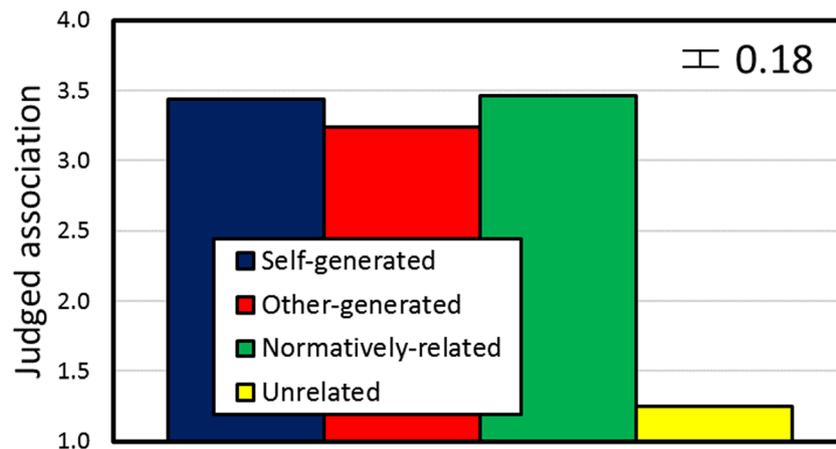


Fig. 4 Judged association of word pairs across the four conditions in Experiment 2

for the previously studied episode more than for the normatively related reminders or reminders generated by others.

The data show that the benefits of reminding are not driven entirely by learners' judged association between the items. Learners rated normatively associated pairs as the most strongly related, yet word pairs in the normative condition were remembered worse than were word pairs in the self-generated condition. As discussed in the Introduction, the effectiveness of self-generated reminders may be driven by idiosyncratic connections to personal experiences (i.e., self-reference) or the reminders' distinctiveness—a comparison we will examine in more depth after the three experiments.

A potential different hypothesis to explain the data is that learners remember the *reminders* (P2s) that they generate better than other reminders. If they remember the reminder that they generated, they could potentially use it to recall the associated P1. For example, during the recall test, if a learner remembers that they generated “library,” they may search their memories for related items (e.g. “book,” “read,” “librarian”) and output related studied items at a greater rate because of strategic recall processes. If this is the case, then the benefits of reminding to P1 would be driven by strategic recall during testing, rather than reminding influences during encoding (Tullis, Benjamin, et al., 2014). If learners were strategically

recalling associated words together, we would expect similarities between the patterns of recall for P1 and P2. However, memory patterns for P2 do not mimic memory patterns for P1. For example, memory for P2s was greatest in the normatively related condition and weaker for P2s that learners generated (even though the pattern for P1s was the reverse). Given the different patterns across memory for P1 and P2, it is likely that memory benefits for reminded P1s are not due solely to memory for P2 and subsequent recall strategies.

Experiment 2

Experiment 1 found mnemonic benefits of self-generated reminders over other conditions in a free recall task. Reminders may match the processes that occur during recall tasks; more specifically, reminders may entail recall of the prior studied word while studying a later related word, just as free recall tests require learners to recall studied items. However, free recall tasks suffer from a lack of control over how learners output items, so learners may strategically search their memories for associates of items that they have recalled (Howard & Kahana, 2002). In Experiment 2, we replicated the procedures of Experiment 1 with a recognition test in order to control

Table 1 Recognition test performance

	P1			P2		
	HR	FAR (rated)	FAR (not rated)	HR	FAR (rated)	FAR (not rated)
Self-generated	0.77 (0.22)	0.31(0.29)		0.71 (0.22)	0.28 (0.26)	
Other-generated	0.74 (0.20)	0.26 (0.27)	0.07 (0.15)	0.71 (0.20)	0.31 (0.25)	0.19 (0.21)
Normatively related	0.72 (0.23)	0.26 (0.22)	0.08 (0.19)	0.75 (0.21)	0.37 (0.24)	0.10 (0.15)
Unrelated	0.73 (0.23)	0.19 (0.19)	0.05 (0.10)	0.74 (0.21)	0.27 (0.25)	0.09 (0.18)

Note. Proportion of hits (recognition ratings of 3 s or 4 s for studied items) and false alarms (recognition ratings of 3 s or 4 s for unstudied items) as a function of condition for P1 and P2. Standard deviations are displayed in parentheses

strategies learners use during testing and minimize output order effects.

Prior research suggests that normative reminders benefit memory only for memory tests involving recall, but not recognition tests (Tullis, Benjamin, et al., 2014). Given that self-generated reminders were more mnemonically beneficial than normative reminders, Experiment 2 is a strong test of whether reminders benefit recognition memory.

Method

Subjects As in Experiment 1, we recruited 45 undergraduate students from the University of Arizona to participate in exchange for partial course credit.

Materials The materials for Experiment 2 consisted of 76 word pairs. Thirty-two pairs were rated by learners during the association phase, and were presented to learners during study (i.e., rated and studied); 32 were rated during the association rating phase, but were not studied (i.e., only rated, not studied); and 12 were not rated during the association phase or studied (i.e., not rated or studied). All unstudied word pairs served as distractors during the recognition test. Word pairs were randomly selected to be studied or not for each participant; further, word pairs were randomly assigned to each of the four conditions as in Experiment 1.

Procedure The procedure of Experiment 2 replicated that of Experiment 1, except for two substantial changes. First, there was only one study/test cycle in Experiment 2. We used one long study list to ensure that recognition was not at ceiling. Second, the memory test involved recognition rather than recall. On the recognition test, participants were presented with a single word, one at a time, and rated their memory for each on a scale of 1 (*I definitely did not study that word*) to 4 (*I definitely did study that word*). The recognition test comprised 32 word pairs that were rated for their association and were studied, 32 words pairs that were rated for their association but were not studied, and 12 word pairs that were not rated or studied. For the 12 word pairs that were neither rated nor studied, four each were in the other-generated, normatively related, and unrelated conditions. Self-generated reminders had to be generated during the preexposure phase, so participants always rated them. During the recognition test, participants were instructed to endorse *only* words seen during the study phase of the experiment and to reject those that they rated during the earlier rating phase but did not study. All targets (P1s) were tested before any reminders (P2s) because we are primarily interested in reminders' effects on the memory for earlier episodes and we wanted to avoid any reminders that could potentially occur among related pairs during testing.

Results

As in Experiment 1, we first analyzed the association judgements that participants made during the preexposure/generation phase, which are shown in Fig. 4. A one-way repeated-measures ANOVA showed that participants' association judgements differed across the four conditions, $F(3, 132) = 774.08, p < .001, \eta_p^2 = 0.95$. Normatively related word pairs were rated as more strongly associated than unrelated pairs, $t(44) = 48.09, p < .001, d = 7.25$. The judged association between normatively related reminders and self-generated reminders did not significantly differ, $t(44) = 0.33, p = .74, d = 0.05$. The judged association between self-generated reminders was significantly greater than other-generated reminders, $t(44) = 3.03, p = .004, d = 0.46$. The self-generated reminders were the same as the normative ones for 15.1% ($SD = 10.0\%$) of the targets.

Next, we examined recognition performance. Hit and false-alarm rates are shown in Table 1.¹ For each condition for each participant, we computed d_a , a signal detection theoretic measure of recognition, and results are shown in Fig. 4. To compute d_a , we used the hit rate for rated and studied items with the false-alarm rate from items that were rated but not studied.² A repeated-measures ANOVA on d_a of P1 showed no significant effect of condition, $F(3, 132) = 1.37, p = .26, \eta_p^2 = 0.03$. Planned comparisons revealed no significant differences between specific conditions: The unrelated condition showed marginally better performance than the normatively related condition, $t(44) = 1.98, p = .05, d = 0.30$, the self-generated condition did not differ from the other-generated condition, $t(44) = 0.02, p = .98, d = 0.004$, and the self-generated condition did not differ from the normative condition, $t(44) = 0.56, p = .58, d = .08$.

We repeated the signal detection analysis on recognition of P2, and the results are shown in Fig. 5. A repeated-measures ANOVA on d_a of P2 showed no significant effect of condition, $F(3, 132) = 1.09, p = .36, \eta_p^2 = 0.02$. Planned comparisons revealed no significant differences between any of the specific conditions: The normatively related condition did not

¹ We computed the hit rate minus the false-alarm rate for each condition, assuming that recognition ratings of 3 s and 4 s were endorsements. We considered endorsements of rated but not studied items to be false alarms and endorsements of rated and studied items to be hits. A repeated-measures ANOVA on hits minus false alarms for P1s showed no significant effect of condition, $F(3, 132) = 1.64, p = .18, \eta_p^2 = 0.04$. No planned comparison between two conditions reached significance ($ps > .05$). The unrelated condition showed marginally higher performance than the normatively related condition, $t(44) = 1.87, p = .07, d = 0.28$.

² We also computed the signal detection theoretic measures of performance between conditions using ratings of studied words as the "old" items and ratings of all items that were *neither* preexposed nor studied as the "new" items. Consistent with the results described in the main text, we found no significant difference in memory across the four conditions, $F(3, 132) = .59, p = .62, \eta_p^2 = .01$. Regardless of the items used as the "old" contrast group, we find no differences in recognition for P1s as a function of the type of P2 studied.

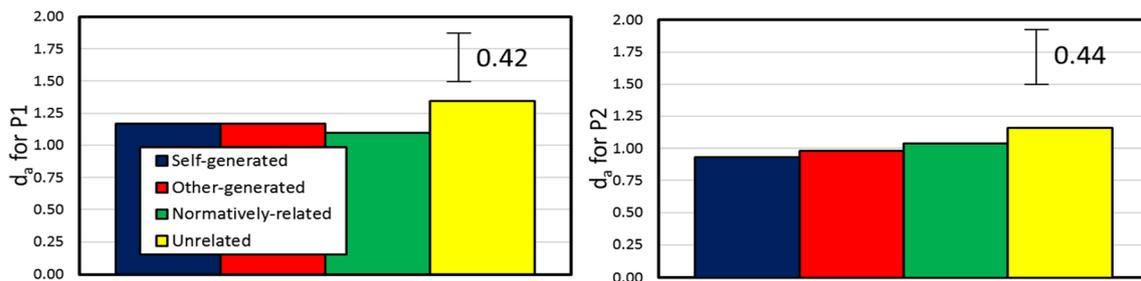


Fig. 5 Signal-detection theoretic measure of recognition performance, d' , for P1s (left panel) and P2s (right panel) in Experiment 2

differ from the unrelated condition, $t(44) = 0.91$, $p = .37$, $d = 0.14$, the self-generated condition did not differ from the other-generated condition, $t(44) = 0.33$, $p = .74$, $d = 0.05$, and the self-generated condition did not differ from the normative condition, $t(44) = 0.80$, $p = .43$, $d = 0.12$.

Discussion

The results from Experiment 2 revealed some similarities and some striking differences from those of Experiment 1. Like Experiment 1, normatively related, other-generated, and self-generated pairs were all judged to be more strongly related than unrelated word pairs. However, unlike Experiment 1, no benefits of reminders appeared on the memory test. In fact, the unrelated condition showed the best recognition of all four conditions. So, while the qualities of the studied words did not shift between Experiments 1 and 2, the mnemonic results did change. Even for very potent reminders (i.e., self-generated reminders), reminders did not boost memory during the recognition test. These results align with those found in the prior literature, suggesting that reminding does not benefit all kinds of memory (Tullis, Benjamin, et al., 2014).

We argue that transfer appropriate processing may determine which tests show mnemonic benefits of reminders. Reminders during study may be more similar to recall tasks than to

recognition tasks because learners *recall* the prior studied information (rather than just recognize it) during a reminding. Therefore, tests that assess recall of content may reveal reminding benefits, while those that do not test recall of the content may not. On the memory test used in Experiment 2, learners do not have to recall the content because each word is presented to them. Instead, the specific recognition test utilized in Experiment 2 may require learners to identify where they have studied the items. To succeed on this task, participants must decipher among three potential sources of memory: everyday life, the preexposure phase, or the study phase. Learners may decipher among these sources by recalling the context of the episode (Johnson, Hashtroudi, & Lindsay, 1993) or by inferring its source from the memory strength (Henkel & Franklin, 1998). These memory processes may differ from recalling the content of episodes, and consequently, reminding benefits may not arise on these tests. However, if reminders change memory because they are a practice test of the earlier episode at the time of the later episode (as suggested by Benjamin & Tullis, 2010), some research suggests that practice tests should bolster memory for the source or context of that memory (Akan, Stanley, & Benjamin, 2018; but see Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010, for more complex results). More broadly, Experiment 2 shows that reminders during study, even with powerful self-generated reminders, did not benefit all types of memory tasks.

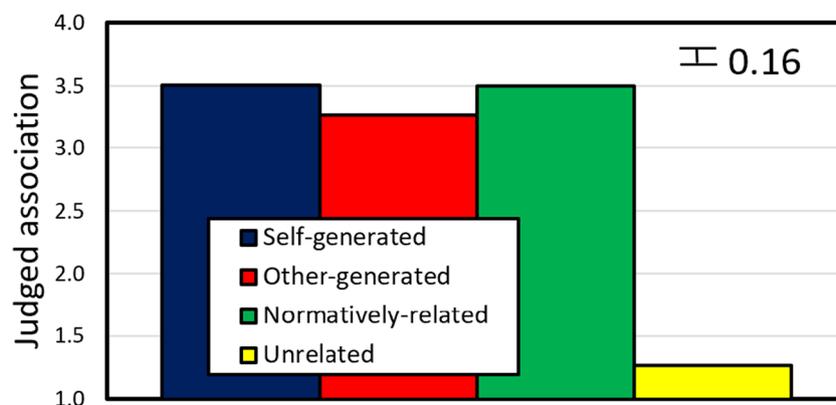


Fig. 6 The judged association of word pairs across the four cue conditions

Experiment 3

Experiment 3 tested whether the lack of mnemonic benefits in recognition result from the control we exerted over test order in Experiment 2 or whether the nature of the memory test (i.e., whether it involves recognition/source identification or retrieval/recall of the content) explains the difference in results between Experiments 1 and 2. In Experiment 3, we utilize an extra-list cued recall test which requires that learners recall the target information rather than recognize it while controlling for output order.

Method

Subjects Tullis, Benjamin, et al. (2014) showed that the mnemonic benefits of reminders in an extra-list cued recall task were smaller than those of free recall, so we computed a new power analysis. A power analysis suggested that 90 participants were needed to detect a small-sized effect ($d = .30$), with 0.8 power and an alpha of 0.05 using a repeated-measures t test. Ninety undergraduate students from introductory educational psychology courses at the University of Arizona participated in exchange for partial course credit.

Materials Thirty-two new word pairs and eight buffer words were selected from the items used in Tullis, Benjamin, et al. (2014; Experiment 3). For each item in the word list, an extra-list cue was selected that was exclusively associated with one of the two items in each pair. For example, if learners studied “navy” and later “army,” “jeep” was the extra-list cue for “army” because “jeep” was unassociated with “navy.” Similarly, “submarine” was the test cue for “navy” because it was not associated with “army.”

Procedure Experiment 3 followed the same general procedures as the previous two experiments, including a preexposure/generation phase, a math task, and a study/test phase. During the study list, participants studied a single word list of 68 items, which included eight pairs in each condition, two primacy untested buffer items, and two recency untested buffer items. The major difference from the prior experiments was that an extra-list cued recall test was utilized to assess memory. During the test, an unstudied extra-list cue was presented with the corresponding initial letter of the target and the correct number of blank spaces for the full target. For example, if the normatively related studied word pair included “square” and “circle,” “square” would be tested with “block-s _ _ _ _.” After all targets (P1s) were tested, the normatively related and unrelated reminders (P2s) were tested. P2s were not tested in the self-generated and other-generated condition because extra-list cues did not exist for targets that were generated by participants.

Results

First, we analyzed the association judgments that participants made during the generation phase, which are shown in Fig. 6. A one-way repeated-measures ANOVA showed that participants’ association judgements differed across the four conditions, $F(3, 267) = 910.16, p < .001, \eta_p^2 = 0.91$. Normatively related word pairs were rated as more strongly associated than unrelated pairs, $t(89) = 44.21, p < .001, d = 4.69$. The judged association between normatively related reminders and self-generated reminders did not significantly differ, $t(89) = 0.03, p = .97, d = .003$. The judged association between self-generated reminders was significantly greater than other-generated reminders, $t(89) = 4.63, p < .001, d = 0.49$. The self-generated reminder was identical to the normative reminder for 32.78% ($SD = 19.86\%$) of the targets.

Next, we examined cued recall of P1s across conditions, which is shown in the left panel of Fig. 7. A repeated-measures ANOVA showed that the cued recall of P1s differed across the four conditions, $F(3, 267) = 5.57, p = .001, \eta_p^2 = 0.06$. Self-generated reminders resulted in better cued recall for P1s than did normatively related reminders, $t(89) = 3.32, p = .001, d = 0.35$. Self-generated reminders did not produce better cued recall for P1s than other-generated reminders, $t(89) = 0.81, p = .42, d = 0.08$.³ Other-generated reminders were more beneficial than normative reminders, $t(89) = 2.46, p = .02, d = .26$. Normatively related reminders did not lead to better cued recall than unrelated reminders, $t(89) = 0.10, p = .92, d = .01$. Finally, we compared cued recall of P2s across conditions, as is shown in the right panel of Fig. 7. Normatively related pairs elicited better cued recall than unrelated pairs, $t(89) = 3.23, p = .002, d = 0.34$.

Discussion

The study procedures of Experiment 3 mirrored those of Experiments 1 and 2, but learners were tested using an extra-list cued recall task. Judged associations across conditions largely replicated that patterns shown in the prior two experiments—namely, self-generated and normatively related reminders were judged to be most strongly related to the target and unrelated reminders were judged to be very weakly related to the targets.

Self-generated reminders boosted memory for prior related episodes more so than normatively related reminders did, as in Experiment 1. One of the explanations of the benefits of self-generated reminders is that participants generate reminders that are tied to their own idiosyncratic knowledge and

³ Given the importance of this comparison for understanding what kind of reminders are most effective, we computed a Bayesian analysis to test how strongly the results support the null or alternative hypotheses. The Bayes factor for this comparison showed substantial evidence in favor of the null (i.e., self- and other-generated reminders being equally effective: $BF = 6.25$).

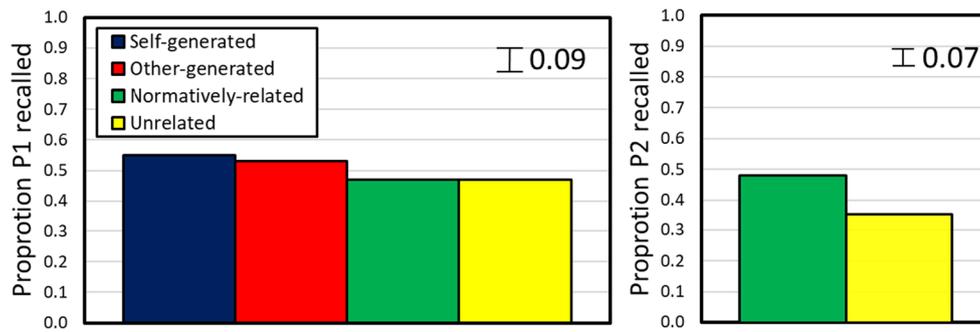


Fig. 7 Cued recall of P1s (left panel) and P2s (right panel) in Experiment 3

experiences. If learners created reminders tied to their own personal experiences, their reminders would be more beneficial for themselves than for others (i.e., self-reference effects). However, data from this experiment refute that hypothesis. Reminders generated by others were just as effective at boosting cued recall of earlier items as reminders generated by oneself. This indicates that self-generated cues in this experiment were not entirely based in idiosyncratic, personal knowledge. Instead, benefits of self-generated reminders may arise because they may be more distinct than normatively related ones; we will explore this hypothesis more fully in the subsequent section.

Unlike prior research using this same extra-list cued recall reminding paradigm (Tullis, Benjamin, et al., 2014), normatively related reminders did not boost cued recall of earlier related episodes more than unrelated reminders did. At least two possibilities may explain why these results did not replicate the prior research. First, in this experiment, learners were preexposed to all pairs during the rating task. Learners may have created episodic connections between unrelated pairs during the judgment of association task and these episodic connections may have supported reminders as much as normatively related reminders. Second, in prior research, only half of the studied words involved reminders (and half were unrelated). Here, three fourths of words involved reminders (only a fourth were unrelated). Competition during the study phase may limit the amount of reminders that are possible. Normatively related pairs may be the weakest reminders and may suffer compared with the stronger reminders in the self-generated and other-generated conditions. Changes in procedures may reduce the benefits of reminding across both

Experiments 1 and 3. To examine whether reminding effects found in these experiments are smaller than those found previously, we compared the mnemonic benefits of reminding in the normatively related condition in Experiment 1 with very similar prior research with normatively related pairs. The benefit of normative over unrelated reminders in Experiment 1 ($d = .51$) was smaller than in very similar prior research (d s of 1.01, 0.84, 0.89: Tullis, Benjamin, et al., 2014). The smaller effects of reminders across Experiments 1 and 3 suggest that the current procedures may reduce the benefits of reminding.

Cue characteristics To determine why self-generated reminders are particularly effective in boosting memory for earlier instances (as shown in Experiments 1 and 3), we examined the characteristics of the reminders used across the conditions. We focused on aspects of reminders that have been shown to be particularly important for cueing memory: normative cue-to-target associative strength and distinctiveness (Tullis & Benjamin, 2015a, 2015b). Self-generated test cues produce high recall when they have high cue-to-target associative strength and distinctiveness. Cue-to-target associative strength has been operationalized as the normative probability that the cue elicits the target during a free association task (Tullis & Benjamin, 2015a); we utilize the South Florida Free Association Norms to provide normative reminder-to-target associative strength (Nelson et al., 1998). Distinctiveness is an indicator of how broadly a reminder is associated; more specifically, we operationalized distinctiveness as the total associative strength from the reminder to every item in the South Florida database (as in Tullis & Fraundorf, 2017). Given that there are no reasons to expect differences between the experiments in the kinds of reminders generated, the results of

Table 2 The associative strength and distinctiveness of the reminders across the four conditions combined across all three experiments

	Self-generated	Other-generated	Normative	Unrelated
Reminder-to-target associative strength	0.15 (0.11)	0.13 (0.10)	0.39 (0.12)	0
Reminder to all targets cumulative strength	0.70 (0.12)	0.69 (0.12)	0.81 (0.04)	0.83 (0.02)

Standard deviations are displayed in parentheses

these analyses combined across experiments are shown in Table 2.

Self-generated and other-generated cues show similar characteristics because self-generated cues were passed to subsequent participants. Compared with normatively related reminders, self-generated reminders showed significantly weaker reminder-to-target associative strength, $t(178) = 30.89$, $p < .001$, $d = 2.32$. However, self-generated reminders showed significantly greater distinctiveness (i.e., they showed smaller cumulative associative strength to all cues within the database) than normatively related reminders, $t(178) = 12.58$, $p < .001$, $d = 0.94$.

General discussion

Self-generated reminders produced greater mnemonic benefits to earlier episodes than normatively related ones on both the free recall and cued recall tests. Normative stimuli allow for well-controlled relationships, but may not harness more distinct relationships. Consequently, normative stimuli, which have been used in all prior research on reminders, may underestimate the impact of reminders on memory. When compared with other-generated reminders, the benefits of self-generated reminders were less consistent. Self-generated reminders boosted memory more than other-generated reminders during free recall, but self-generated reminders and other-generated reminders produced equal benefits when we controlled output order with an extra-list cued recall test. The benefits of self-generated reminders over other-generated reminders on the free recall test may be driven by high free recall of self-generated reminders (P2). Because learners generated the reminders in Experiment 1, they are likely to recall a high proportion of those reminders (i.e., generation effects; Slamecka & Graf, 1978). After learners recall the reminder (P2) that they generated, they may use bootstrapping strategies to recall the associated target (P1) during the recall test. When controlling for output strategy use during recall (as in Experiment 3), self-generated reminders provided no benefits compared to reminders generated by others. The results suggest that self-generated and other-generated reminders boost recall to the same extent when output strategies are controlled.

Self-generated reminders more effectively elicit reminders than do normative ones, but the results imply that self-generated reminders do not derive their effectiveness from their idiosyncratic connections to one's personal experiences. If self-generated reminders derived their effectiveness from connections to idiosyncratic personal knowledge (i.e., self-reference), reminders generated by others would not be as helpful as reminders generated by oneself (Symons & Johnson, 1997). One limitation of this explanation is that our sample may be sufficiently homogeneous that self-generated cues largely overlap with other participants' experiences. If so,

self-generated reminders would be equally effective for others. To test this idea, we calculated the proportion of self-generated cues that overlapped among participants. Across the three experiments, the percentage of participants who generated the same cue for each target averaged 17% ($SD = 8\%$), indicating significant variety among the reminders that participants generated. Even though reminders varied among learners, other-generated reminders boosted memory for targets as much as self-generated reminders in Experiment 3. This, again, suggests that self-reference effects (or the idiosyncrasies of self-generated reminders) do not cause the benefits seen in self-generated reminders.

Further, the data suggest that the benefits of self-generated reminders are not driven entirely by generation effects. If generation, in and of itself, underlay the mnemonic effects of self-generated reminders, self-generated reminders would be more effective than other-generated reminders. Learners' recall of P1 benefits as much from studying others' reminders as they do from studying their own reminders. Second, our primary conclusions derive from comparisons across memory for P1s, which are words that learners did not generate. The mnemonic benefits of generation may strengthen cue-to-target processing (e.g., McDaniel, Waddill, & Einstein, 1988), but, if generation strengthened reminder-to-target processing in our experiments, we would likely see bigger benefits for self-generated reminders than for other-generated reminders. Further, extensive research suggests that generation boosts memory across both recall and recognition tasks (for a meta-analysis, see Bertsch, Pesta, Wiscott, & McDaniel, 2007). We find no benefits of generation on the recognition test in Experiment 2. For these reasons, we believe that generating reminders (i.e., P2s) did not produce mnemonic benefits for related P1s.

We also argue that reminder-to-target associative strength does not exclusively determine the effectiveness of reminder. Participants reported greater reminder-to-target associative strength for normatively related reminders than other-generated ones, but other-generated reminders boosted memory more. The data consistently show that the strength of the association between the reminder and target does not exclusively determine the effectiveness of the reminder. Further, learners may judge the strength of association based upon the bidirectional association from target to reminder and from reminder to target (e.g., Koriat & Bjork, 2005). Reminding theory posits that the backward associative strength (from reminder to target) is more crucial for effective reminding than forward associative strength (from target towards reminder), but no research has specifically examined this question. Future research can examine whether the direction of this relationship matters for the effectiveness of a reminder.

We finally examined how the distinctiveness of reminders relates to efficacy of reminders, given that connections to one's personal idiosyncratic experiences, generation, and reminder-to-target associative strength did not entirely

determine the effectiveness of reminders. Distinctiveness indicates the breadth of associations from a reminder and may play an important role in guiding learners to a single prior episode. When learners generate cues for specific target words, learners output distinctive cues and distinctiveness is a prime factor in the cues' effectiveness (Tullis & Fraundorf, 2017). The benefits of self-generated and other-generated reminders, then, may be partially driven by the distinctiveness of those reminders. In these experiments, normatively related reminders were more broadly related to potential targets (i.e., less distinct) than reminders generated by individuals. Self-generated and other-generated reminders may derive effectiveness, not exclusively through reminder-to-target associative strength but also through their distinctiveness. In other words, another difference between conditions is the number of participants who have generated the reminders. In the normative condition, reminders are based upon a broad participant pool and reflect shared, consistent knowledge across learners. In the self-generated and other-generated conditions, reminders are based upon a single participant. There is great variability in the kinds of reminders that learners generate. The data show that this variability may be helpful because it allows for distinctive reminders that point to fewer potential target items.

One difference between the current procedures and prior reminding procedures is the inclusion of the preexposure/generation phase. In order to test the effectiveness of self-generated reminders, we must elicit self-generated reminders from learners before study. We equated the impact of the preexposure/generation phase of the experiment by incorporating all conditions in an association rating task. The preexposure phase presented words from all four conditions to participants and ensured that participants processed those pairs as similarly as possible. Yet the preexposure phase could potentially impact later memory. For example, studying a P1 could potentially remind learners of related P2s from the preexposure phase (however, we make no strong conclusions about memory for P2s). To reduce the impact of the preexposure phase on reminding, we separated the preexposure phase from the study/test phase as much as possible within our hour-long experimental timeslot and made no explicit mention of the connection of the preexposure phase to the study/test phase. Again, a significant limitation of these data is that learners were preexposed to all studied items during the generation/preexposure phase. If the conditions caused learners to differentially encode P1s between the conditions during the preexposure/generation phase, this would confound the potency of reminders during the later study session. For example, generating a P2 (a reminder) could potentially change long-term memory for P1 and mnemonic advantages to P1 may arise due to this phase rather than due to reminders.⁴ This preexposure phase may potentially also induce mnemonic benefits across unrelated reminders later

during study and obscure differences between conditions. Future research may try to further reduce the impact of the preexposure phase on later studying using longer delays or more items during the preexposure phase.

Notably, our self-generation instructions prompted learners to produce words that “made them think of the target word” rather than explicitly asking participants to generate “reminders” for the target words. We avoided including “reminding” in the instructions to minimize demand characteristics that could unfairly benefit the self-generated condition. Further, we believed that asking how strongly a word “reminded” learners of a different word may have confused learners. Whether participants would output different kinds of reminders with different instructions could be explored in future research. Modifying the instructions to prompt learners to output “reminders” could potentially bolster the efficacy of self-generated reminders.

Our research questions center exclusively on how reminders impact memory for earlier information; more specifically, how do different kinds of P2s retroactively impact memory for earlier studied P1s? We cannot draw conclusions about the proactive influence of P1s on memory for P2s across conditions for multiple reasons. First, we cannot control the stimulus characteristics across P2s, which are drawn from norms in two conditions and generated by participants in the other two conditions. Differences in stimulus characteristics preclude meaningful comparison of memory for P2s across conditions. Second, in Experiments 2 and 3, we test memory for P1s before we test memory for P2s. We control testing order in this manner because testing one item may affect memory for related items and we prioritize measuring memory for P1 cleanly. Given that we test P2s only after testing P1s, we do not make strong conclusions about the impact of P1s on memory for later P2s. P1s may influence memory for later P2s, but that question is beyond the scope of the current experiments. Future research should cleanly test the proactive effects of a related P1 on later P2s (see Wahlheim & Jacoby, 2013, for an example).

Despite the efficacy of self-generated reminders in Experiments 1 and 3, the recognition test showed no mnemonic benefits of self-generated reminders. These results align with prior research, which suggests that reminders boost future recall of instances, but do not enhance future recognition (Tullis, Benjamin, et al., 2014). Reminders may involve practice recall of the prior related word during study of a later word; practice recall is typically most beneficial for future recall tests (Chan & McDermott, 2007). The consequences of reminders on memory may be driven by the boost to recallability for the earlier content that practice recall provides.

⁴ As we have argued earlier in the Discussion section, we believe that the data point against this explanation because self-generated and other-generated targets are remembered equally well in the third experiment.

Other theories of reminding suggest that studying related episodes benefits memory because learners consciously rehearse earlier episodes during later related information (McKinley, Ross, & Benjamin, 2019; Rundus, 1971). The current results hint that the benefits of reminders are not driven by extra conscious rehearsal of the target during study of the reminder because extra rehearsal would enhance recognition performance for those studied items. We find no benefit of reminding on recognition. Finally, other theories of reminding suggest that reminders benefit memory through spreading activation mechanisms (e.g., Collins & Loftus, 1975). In contrast to those theories, our results suggest that strongest associated items were not the most effective reminders. The results of these three experiments, then, may be most consistent with the idea that reminders are beneficial because they involve practice recall of prior episodes.

People frequently report generating reminders in their everyday lives, including making to-do lists, placing items in unusual places, and creating calendar entries. The capability of people to generate effective reminders may be crucial to succeeding in a very busy world. For example, the ability of a calendar entry to trigger the appropriate memory is required to instigate appropriate actions and achieve important goals. In our studies, personally generated reminders yielded greater mnemonic benefits for earlier target information than normative reminders. The benefits of self-generated reminders may be driven by their distinctiveness, a characteristic that people value when generating memory cues (Tullis & Benjamin, 2015b). The ability to create effective reminders adds to the long list of metacognitive control that learners can effectively exercise over their memories (e.g., Finley, Tullis, & Benjamin, 2009), including successfully choosing study strategies (Tullis, Fiechter, & Benjamin, 2018) and judiciously allocating study time across items (Tullis & Benjamin, 2011). However, the ability to generate effective reminders is unique because learners often fail to show effective control over their future retrieval (Finley & Benjamin, 2020). Learners in our experiments generated reminders for their future mindsets that effectively provoked reminders and benefited memory. Understanding when and how learners utilize personal knowledge of their own cognitive states to control their retrieval environments can shape whether and how we should provide support to learners who are exercising metacognitive control over their memories.

The data and materials for all experiments are available (https://osf.io/w9m3s/?view_only=3d50437869dd4bf7b7555d5ee797b392).

Funding This research was supported by an NSF CAREER grant (Award 1847007) to the PI entitled “That Reminds Me: The Causes and Consequences of Reminders.”

Appendix

Between-participant correlations between conditions across the three experiments.

Table 3 Correlations among the judged associations in Experiment 1

	Other-Gen	Normative	Unrelated
Self-Gen	.18	.13	.33
Other-Gen		.18	.35
Normative			.31

Table 4 Correlations among the proportion of free recall for P1s in Experiment 1

	Other-Gen	Normative	Unrelated
Self-Gen	.07	.21	.26
Other-Gen		.29	.18
Normative			.06

Table 5 Correlations among the proportion of free recall for P2s in Experiment 1

	Other-Gen	Normative	Unrelated
Self-Gen	.22	-0.06	-0.06
Other-Gen		.27	.09
Normative			.25

Table 6 Correlations among the judged associations in Experiment 2

	Other-Gen	Normative	Unrelated
Self-Gen	0.20	0.63	0.34
Other-Gen		0.40	0.01
Normative			0.27

Table 7 Correlations among the d_a s of recognition for P1s in Experiment 2

	Other-Gen	Normative	Unrelated
Self-Gen	0.51	0.45	0.44
Other-Gen		0.52	0.40
Normative			0.52

Table 8 Correlations among the d_a s of recognition for P2s in Experiment 2

	Other-Gen	Normative	Unrelated
Self-Gen	0.19	0.40	0.50
Other-Gen		0.53	0.32
Normative			0.42

Table 9 Correlations among the judged associations in Experiment 3

	Other-Gen	Normative	Unrelated
Self-Gen	0.42	0.50	0.01
Other-Gen		0.47	0.21
Normative			0.11

Table 10 Correlations among the proportion cued recall for P1s in Experiment 3

	Other-Gen	Normative	Unrelated
Self-Gen	0.47	0.45	0.51
Other-Gen		0.50	0.36
Normative			0.37

Table 11 Correlations among the proportion cued recall for P2s in Experiment 3

	Unrelated
Normative	0.46

References

- Akan, M., Stanley, S. E., & Benjamin, A. S. (2018). Testing enhances memory for context. *Journal of Memory and Language*, *103*, 19–27.
- Bäckman, L., & Mäntylä, T. (1988). Effectiveness of self-generated cues in younger and older adults: The role of retention interval. *The International Journal of Aging and Human Development*, *26*(4), 241–248.
- Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 153.
- Benjamin, A. S., & Ross, B. H. (2011). The causes and consequences of reminding. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 71–87). New York: Psychology.
- Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247.
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210.
- Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory*, *18*, 385–393.
- Bruce, D., & Weaver, G. E. (1973). Retroactive facilitation in short-term retention of minimally learned paired associates. *Journal of Experimental Psychology*, *100*(1), 9.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380.
- Chan, J. C., & McDermott, K. B. (2007). The testing effect in recognition memory: a dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 431.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.
- 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.
- Finley, J. R., & Benjamin, A. S. (2020). *Metacognitive control of transfer-appropriate processing*. Manuscript in preparation.
- Finley, J. R., Tullis, J. G., & Benjamin, A. S. (2009). Metacognitive control of learning and remembering. In M. S. Khine & I. M. Saleh (Eds.) *New science of learning: Cognition, computers and collaboration in education*. New York: Springer.
- Fraundorf, S. H., Diaz, M. I., Finley, J. R., Lewis, M. L., Tooley, K. M., Isaacs, A. M., ... Brehm, L. (2014). CogToolbox for MATLAB. Retrieved from <http://www.scottfraundorf.com/cogtoolbox.html>. Accessed 5 Oct 2018.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.
- Harris, J. E. (1980). Memory aids people use: Two interview studies. *Memory & Cognition*, *8*(1), 31–38.
- Henkel, L. A., & Franklin, N. (1998). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition*, *26*(4), 659–673.
- Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, *32*(2), 336–350.
- Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, *38*(1), 102–115.
- Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior*, *12*(2), 229–238.
- Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 31–40.
- Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*(1), 85–98.
- Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). New York: Oxford University Press.
- Hunt, R. R., & Smith, R. E. (1996). Accessing the particular from the general: The power of distinctiveness in the context of organization. *Memory & Cognition*, *24*(2), 217–225.

- Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive reminders in recency judgments and cued recall. *Memory & Cognition*, 41(5), 625–637.
- Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1282.
- Jacoby, L. L., Wahlheim, C. N., & Yonelinas, A. P. (2013). The role of detection and recollection of change in list discrimination. *Memory & Cognition*, 41(5), 638–649.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3.
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 187–194.
- Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of the National Academy of Sciences of the United States of America*, 97(1), 506–511.
- Lofus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490.
- Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12, 66–71.
- Mäntylä, T., & Nilsson, L. G. (1983). Are my cues better than your cues? Uniqueness and reconstruction as prerequisites for optimal recall of verbal materials. *Scandinavian Journal of Psychology*, 24(1), 303–312.
- Mäntylä, T., & Nilsson, L. G. (1988). Cue distinctiveness and forgetting: Effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 502.
- McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, 27(5), 521–536.
- McKinley, G., Ross, B. H., & Benjamin, A. S. (2019). The role of retrieval during study: Evidence of reminding from self-paced study time. *Memory & Cognition*, 47, 877–892.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms*. Retrieved from <http://www.usf.edu/FreeAssociation/>. Accessed 6 Sept 2018.
- Osterberg, L., & Blaschke, T. (2005). Adherence to medication. *The New England Journal of Medicine*, 353, 487–497.
- Reeves, L., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, 115(3), 381.
- Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16(3), 371–416.
- Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 629.
- Ross, B. H. (1996). Category learning as problem solving. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 35, pp. 165–192). Cambridge: Academic.
- Ross, B. H., & Bradshaw, G. L. (1994). Encoding effects of reminders. *Memory & Cognition*, 22(5), 591–605.
- Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 42.
- Ross, B. H., Perkins, S. J., & Tenpenny, P. L. (1990). Reminding-based category learning. *Cognitive Psychology*, 22(4), 460–492.
- Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63–77.
- Slamecka, N.J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 592–604.
- Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, 121(3), 371.
- Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, 64(2), 109–118.
- Tullis, J. G., & Benjamin, A. S. (2015a). Cue generation: How learners flexibly support future retrieval. *Memory & Cognition*, 43(6), 922–938.
- Tullis, J. G., & Benjamin, A. S. (2015b). Cueing others' memories. *Memory & Cognition*, 43(4), 634–646.
- Tullis, J. G., & Benjamin, A. S. (2020). *Reminders influence memory for context*. Manuscript in preparation.
- Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General*, 143(4), 1526.
- Tullis, J. G., Braverman, M., Ross, B. H., & Benjamin, A. S. (2014). Reminders influence the interpretation of ambiguous stimuli. *Psychonomic Bulletin & Review*, 21(1), 107–113.
- Tullis, J. G., Fiechter, J. L., & Benjamin, A. S. (2018). The efficacy of learners' testing choices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 540.
- Tullis, J. G., & Finley, J. R. (2018). Self-generated memory cues: effective tools for learning, training, and remembering. *Policy Insights from the Behavioral and Brain Sciences*, 5(2), 179–186.
- Tullis, J. G., & Fraundorf, S. H. (2017). Predicting others' memory performance: The accuracy and bases of social metacognition. *Journal of Memory and Language*, 95, 124–137.
- Tullis, J. G., & Goldstone, R. L. (2016). Comparison versus reminding. *Cognitive Research: Principles and Implications*, 1(1), 20.
- Tzeng, O. J., & Cotton, B. (1980). A study-phase retrieval model of temporal coding. *Journal of Experimental Psychology: Human Learning and Memory*, 6(6), 705.
- Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall. *Memory & Cognition*, 26(1), 108–120.
- Wahlheim C. N., & Jacoby L. L. (2013). Remembering change: The critical role of recursive reminders in proactive effects of memory. *Memory & Cognition*, 41, 1–15.
- Watkins, O. C., & Watkins, M. J. (1976). Cue-overload theory and the method of interpolated attributes. *Bulletin of the Psychonomic Society*, 7(3), 289–291.
- Winograd, E., & Soloway, R. M. (1985). Reminding as a basis for temporal judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(2), 262.

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