What characteristics make self-generated memory cues effective over time?

Jonathan G. Tullis & Jason R. Finley

MEMORY
https://doi.org/10.1080/09658211.2021.1979585

What characteristics make self-generated memory cues effective over time?

Jonathan G. Tullis [a] and Jason R. Finley [b]

[a]Educational Psychology, University of Arizona, Tucson, AZ, USA; [b]Psychology, Southern Illinois University Edwardsville, Edwardsville, IL, USA

ABSTRACT

From writing to-do lists to creating mnemonic devices in school, people frequently generate cues to help them remember information. Creating memory cues is a vital aspect of metacognition and allows learners to somewhat control their retrieval circumstances. Across three experiments, we tested the extent to which self-generated memory cues fail at long retention intervals because they are based in fleeting mental states. Participants studied target words and generated mnemonic cues for themselves or for others. Cues intended for others showed greater cue-to-target associative strength, were less distinctive, and were less idiosyncratic (more common) than cues intended for oneself. However, the effectiveness of the cues in supporting recall did not differ by intended recipient at medium (~3 days) or long (~1 year) retention intervals. In the third experiment, we directly tested the stability of self-generated cues for oneself (compared to cues for others, descriptions of the target, and focused descriptions) by asking participants to generate cues twice for the same targets across a delay of 3 weeks. Cues intended for others were more stable than all other cues, but the stability of the cues did not affect long term retention. Implications for effective cue generation are discussed.

ARTICLE HISTORY
Received 5 February 2021
Accepted 8 September 2021

KEYWORDS
Metacognition; metacognitive control; cue generation; stability bias; perspective taking

Have you ever written a note to yourself, and then later struggled to figure out what you meant? Creating mnemonic cues to support one's memory is a crucial aspect of metacognition. People generate memory cues regularly, including naming computer files so that one remembers their contents, writing to-do lists so that one remembers to complete important tasks, and creating mnemonics so that one remembers the names of the Great Lakes for an upcoming test. In fact, 73% of people report using the first letter of to-be-remembered items to create a more memorable structure for information (e.g., ROY G. BIV for the colours in the rainbow) and 57% of people report using rhymes to help them remember (e.g., *i* before *e* except after *c*: Harris, 1980). Students of all ages report creating mnemonics to help remember classroom information (Tullis & Maddox, 2020; Van Etten et al., 1997). These self-generated memory cues support recall more effectively than cues generated by others across a variety of tasks (Bellezza & Poplawsky, 1974; Jamieson & Schimpf, 1980; Kuo & Hooper, 2004; Saber & Johnson, 2008; for review, see Tullis & Finley, 2018). Yet, external mnemonic cues sometimes fail. Across three experiments, we examine whether self-generated mnemonic cues fail because they are generated in fleeting, unstable mental states which do not ultimately match mental states at long retention intervals.

Cue generation is a vital aspect of metacognitive control that people exercise over their memories. The control that people exercise over their memories during encoding and retrieval may have greater impact on their recall than individual differences in mnemonic abilities (Benjamin, 2008). Learners make many effective choices during encoding, including choosing which items to restudy (Kornell & Metcalfe, 2006; Tullis & Benjamin, 2012), allocating study time (Tullis & Benjamin, 2011), scheduling study (Benjamin & Bird, 2006), choosing encoding strategy (Finley & Benjamin, 2012), and even choosing retrieval practice over restudy (Tullis et al., 2018). Learners may be able to choose effective encoding practices because they have privileged access to their own idiosyncratic mental states during encoding and can therefore base choices on this privileged knowledge that others do not have (Lovelace, 1984; Underwood, 1966). In other words, if learners believe that they have encoded an item poorly, they can choose to restudy that particular item. Cue generation is a particularly interesting example of metacognitive control over memory because learners need to anticipate what cue will be most helpful during later retrieval. In essence, they must lay down a trail of

CONTACT Jonathan G. Tullis tullis@arizona.edu Department of Educational Psychology, 1430 E. Second St., Tucson, AZ 85721, USA
Supplemental data for this article can be accessed at https://10.1080/09658211.2021.1979585

breadcrumbs for their future selves to follow. As described below, anticipating one's future mental state is a difficult task. Here, we examine whether failure to anticipate future mental states impairs cue generation.

Typically, learners generate effective cues that trigger their recall. In a classic laboratory experiment, participants remembered 91% of words from a very long list when they were prompted with self-generated cues (Mäntylä, 1986). Self-generated cues benefit memory for a variety of target information, including simple words (Mäntylä, 1986; Zhang & Tullis, 2021; Mantyla & Nilsson, 1983), foreign language vocabulary (Atkinson & Raugh, 1975), and complex science concepts (Levin & Levin, 1990; Richmond et al., 2011; Tullis & Qui, 2021). Self-generated cues are typically more effective than cues generated by peers (Tullis & Benjamin, 2015b), produced by experts (Bloom & Lamkin, 2006), or randomly selected from an experimenter-curated list (Finley & Benjamin, 2019). Prior research argues that self-generated cues can be effective because learners choose cues that are tailored to their specific memory needs (Tullis & Benjamin, 2015b). Learners have special access to their own idiosyncratic cognitive states and prior experiences (Lovelace, 1984; Tullis & Fraundorf, 2017; Underwood, 1966), and this private knowledge about their own mental states may allow them to output uniquely effective cues (Kuo & Hooper, 2004; Symons & Johnson, 1997; Tullis & Finley, 2018). Specifically, when learners generate mnemonic cues, they output distinctive cues (i.e., cues that narrowly point to a small set of potential targets; Tullis & Fraundorf, 2017). The distinctiveness of the cue narrows the search space for the target at the time of retrieval.

Despite the broad effectiveness of self-generated mnemonic cues, self-generated mnemonic cues can fail, especially at long retention intervals. Self-generated memory cues may fail because learners struggle to anticipate future mental states. Accurately anticipating future mental states is crucial to effective cue generation because the conditions under which learners generate cues differ from those under which they attempt retrieval. For example, when generating cues, the target is present (and the cue is present once the learner generates it); but, during retrieval, only the cue is present. So, the generating conditions do not match the retrieval conditions. Beyond differences in what is present during generation and what is present at retrieval, learners' mental states naturally fluctuate across time (Estes, 1955). Retrieval depends upon the overlap between encoding and retrieval contexts, so successful mnemonic cues generated during encoding must partially match the learner's cognitive state at the time of retrieval (e.g., Raaijmakers & Shiffrin, 1980; Ryskin et al., 2015). The concepts that are activated now are unlikely to be present or activated in a year.

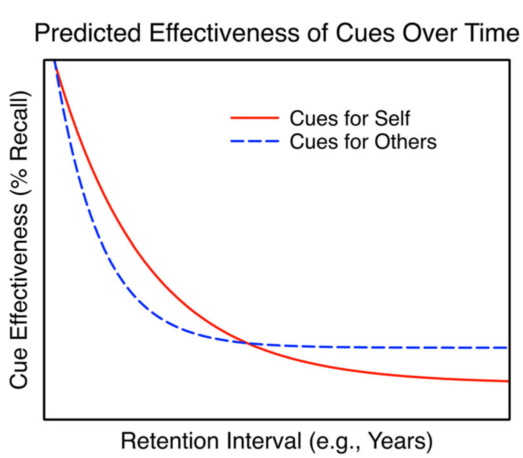

Figure 1. Predicted effectiveness of cues over time (hypothetical results).

Anticipating and accounting for the differences between current and future mental states may be difficult (see Koriat & Bjork, 2005), especially as learners often show a stability bias, in which they make judgments based upon their immediate subjective experience (cognitive egocentrism) rather than on beliefs about future events (Kornell et al., 2011; Kornell & Bjork, 2009). Generating effective memory cues requires theory of mind for ourselves in the future, and people tend not to recognise the extent to which they become different people over time (cf. the "end of history illusion," Quoidbach et al., 2013). The longer the retention interval, the more people change. Body and mind age, new experiences accumulate, new perspectives develop, language changes, social and occupational circumstances change. Things that may seem obvious now can become obscured with time. Making accurate choices about the future can necessitate divorcing those choices from current mental states, which is difficult to do during encoding (Kornell et al., 2011). Further, even when people have been told about the future circumstances (e.g., that test conditions will differ from encoding conditions), people struggle to take the perspective of themselves in the future, and thus make inaccurate predictions of their memories (Koriat & Bjork, 2005).

Across three experiments, we tested whether and how learners generate cues that are stable and support memory across long retention intervals. One plausible mechanism for ensuring stable and long-lasting cues is to prompt participants to generate cues for others. Prior research suggests that when learners generate mnemonic cues to help other people's memories, the generated cues show stronger cue-to-target associative strength and lesser idiosyncracies than cues intended for oneself (Tullis & Benjamin, 2015b). While cues for oneself may be based on idiosyncratic and temporary mental activation, cues for others may be based more in stable, normative knowledge that does not shift much across time. Across three experiments, we tested whether cues intended for others elicit recall more than cues intended for oneself at extended retention intervals. We predicted that distinctive and idiosyncratic self-generated memory cues will be more effective at supporting recall in the near future, but eventually people and their circumstances will change enough that cues generated for others will become more effective. This hypothetical pattern is illustrated in Figure 1. We specifically assess three characteristics of the generated mnemonic cues across the experiments, as in prior research (Tullis & Benjamin, 2015a, 2015b; Tullis & Fraundorf, 2017). First, we measured the cue-to-target associative strength from the South Florida Free Association Norms (Nelson et al., 1998). Cue-to-target associative strength indicates how strongly the cue normatively points towards the appropriate target; cues for others typically have greater normative cue-to-target associative strength. Second, we measure the distinctiveness of each cue by assessing how many total targets are associated with the cue in the South Florida Free Association Norms. Fewer total associates narrows the search space during the time of retrieval and may be one of the most important factors that drives recall (Hunt & Smith, 1996; Nairne, 2002). Cues for others typically have less distinctiveness than cues for oneself. Finally, we measure how common each cue is so that we can assess how idiosyncratic the cue is across learners. Cue commonality reflects the proportion of participants who supplied a given cue for each target and is typically greater for cues for others than for oneself (Tullis & Benjamin, 2015b). These prior studies suggest that cues for others tap into more normative and common knowledge, while cues for oneself are more idiosyncratic and distinctive.

Ultimately, we sought to answer three research questions in our three experiments. First, do people generate different cues for their own memories and others' memories, as suggested in prior research (Tullis & Benjamin, 2015b)? Second, do cues generated for others support learners' recall at long retention intervals better than cues intended for oneself because their connection to the target information is more stable? In other words, if cues intended for oneself are based in fleeting mental states, they may become less effective than cues for others at long retention intervals (Figure 1). Finally, do the characteristics of effective cues change with retention interval? More specifically, are effective cues at long retention interval tied into common, stable, semantic knowledge, while effective cues at short retention intervals tied in to more idiosyncratic and distinctive knowledge?

Experiment 1

In Experiment 1, participants generated cues for themselves or for peers and we tested the effectiveness of those cues at two different retention intervals. If participants generated cues for themselves in fleeting, idiosyncratic mental states, cues for themselves should become much less effective at longer retention intervals. If cues for others are based upon normative, enduring knowledge, then cues for others should support recall better than cues for self at longer lags.

Method

Participants. We sampled the entire class of students of Fontbonne University enrolled in an intermediate level psychology methods course. Sixty-nine students generated cues during an initial experimental session, including 57 females, 11 males, and 1 unreported. The mean age of students was 22.27 ($SD = 5.54$). Forty-two students completed the first test which happened several days after the cue generation phase ($M = 5$ days, $SD = 5$, $Mdn = 3$, range: 1–22), and 28 students completed the final test which happened approximately one year after the cue generation phase ($M = 385$ days, $SD = 18$, $Mdn = 379$, range: 359–421).

Design. This experiment featured two within-subjects independent variables: retention interval (~3 days for test 1 vs. ~1 year for test 2) and intended recipient of the cues that participants generated (self vs. other). The dependent variables measured were the cues that participants generated and their performance on the cued recall tests.

Materials. One hundred and sixty targets were collected from the University of South Florida Free Association Norms (Nelson et al., 1998). We gathered a wide variety of potential target items.

Procedure. Participants completed the experiment online. In this and subsequent experiments, participants first completed their informed consent. Next, we instructed participants to remember a list of target words for an upcoming memory test. We told them that they would generate a cue to help them remember each target. Participants were instructed that they would get a cue back during the test and would have to recall the corresponding target word. Cues could be any single, English word that was not a form of the target (i.e., a plural or a misspelling of the target). Participants were instructed that they would sometimes generate cues to help their own memory and sometimes they would generate cues to help a peer's memory. Peers were described as their classmates who were also participating in this experiment. Half of target items were assigned to the self condition and half were assigned to the peer condition. The target was displayed on the right hand side of the screen and participants had to type their cue into an empty response box on the left hand side of the screen. The self and peer condition were differentiated on the screen by placing the target and response box for one condition consistently towards the top of the screen and the other condition towards the bottom of the screen; the positioning of conditions was counterbalanced across participants. Half of the target words were randomly assigned to the self condition and half were assigned to the other condition.

Two days after generating the cues, participants were contacted to complete the second portion of the experiment. On the second day, participants were shown 80 cues that they generated (40 in each condition) and typed in the corresponding target item. Cues were presented one at a time in an entirely random order. Participants were emailed every 2 days (up to 5 emails) to complete this first test. Finally, starting 1 year after they completed the initial test, we emailed participants to ask them to complete the final test. We emailed them every 2 days up to 5 emails to ask to finish the experiment. On this last testing day, participants were tested with the remaining 80 untested cues (40 from the self condition and 40 from the other condition).

Results

The data and syntax for this and the subsequent experiments are available (Tullis & Finley, 2021, February 4) here: https://osf.io/z4f6c/?view_only=a0fc8a6f83e149cca31cee2b3e855274.

Cues for self and others. We first examined whether participants differentiated between the cues intended for themselves versus others. To do so, we compared the cue characteristics across all participants (the same patterns appear if we include only the participants who completed both tests). Data are shown in Table 1 and indicate that participants generated different kinds of cues for self and for others. More specifically, when generating cues for themselves, participants created cues that were more distinct (i.e., had fewer associates), were more unique (i.e., showed lesser cue commonality), and had weaker cue-to-target associative strength than when generating cues for others.[1]

Cued Recall. Next we examined how effectively cues supported recall of target words as a function of test number (1 vs. 2) and intended recipient of cues (self vs. other). Results are shown in Figure 2. In order to take advantage of the large number of observations per participant, we analysed the data using a linear mixed effect model with a logit link function to model the log odds of correct recall of each target word. The model included fixed effects of test number, intended recipient, and their interaction, as well as crossed random intercepts for participants and target words. We coded fixed effects of test number and intended recipient as −0.5 and +0.5 to obtain estimates of the main effects analogous to those from an ANOVA. For this analysis we used the glmer() function of the lme4 package for R (Bates et al., 2015). Syntax is available in supplemental materials. Only participants that completed both test sessions were included in this analysis. The model was:

$$\begin{aligned}\text{logit}(\text{P}(Recall_{ij})) = (\beta_0 + u_{0i} + v_{0j}) + \beta_1 TestNumber_{ij} \\ + \beta_2 IntendedRecipient_{ij} + \beta_3 Interaction_{ij}\end{aligned} \tag{1}$$

where i is the participant index, j is the target word index, β_0 is the baseline intercept, u_{0i} is the random intercept for participants, v_{0j} is the random intercept for target words, and $\beta_{1\text{-}3}$ are fixed-effect coefficients (i.e., slopes) for test number, intended recipient, and the interaction between test number and intended recipient, respectively. Table 2 shows the results.

There was only a statistically significant effect of test number (i.e., retention interval), such that performance was lower after 1 year versus 3 days. Neither the intended recipient nor the interaction were statistically significant. Our prediction was that recall would be better when prompted with one's own cues after a few days, but would be better when prompted with the cues made for others after a year. We did not find that pattern.

Relation of cue characteristics to recall. Finally, we examined how cue characteristics related to their effectiveness. To do so, we computed a linear mixed model predicting recall from test number (1 or 2), the three cue characteristics, and the interactions between the cue characteristics and test number. We centred cue characteristic variables before adding them to the model. As shown

Table 1. Means and standard deviations in Experiments 1and 2 of cue characteristics for Cue-target pairs.

Experiment	Cue characteristic	For others	For self	
Experiment 1	Cue-to-target associative strength	.07 (.03)	.06 (.02)	$t(68) = 4.17$, $p < .001$, $d = .51$
	Number of associates	9.87 (1.34)	9.51(1.61)	$t(68) = 2.57$, $p = .01$, $d = .31$
	Cue commonality	.14 (.03)	.13 (.03)	$t(68) = 4.14$, $p < .001$, $d = .50$
Experiment 2	Cue-to-target associative strength	.08 (.03)	.06 (.03)	$t(54) = 5.31$, $p < .001$, $d = .73$
	Number of associates	9.68 (1.64)	9.20(2.15)	$t(54) = 2.18$, $p = .03$, $d = .30$
	Cue commonality	.15 (.03)	.13 (.03)	$t(54) = 3.34$, $p = .002$, $d = .46$

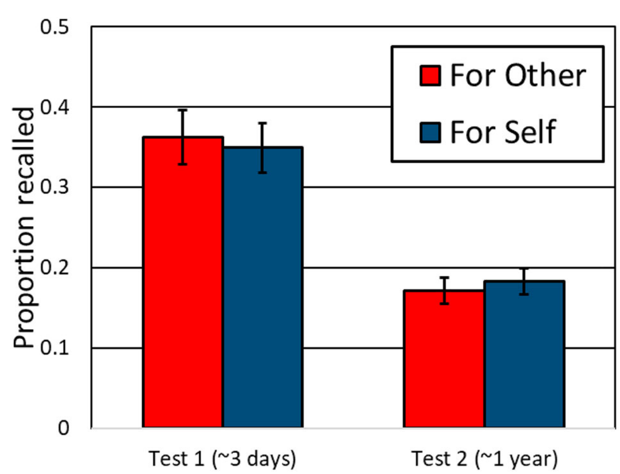

Figure 2. Mean proportion recall as a function of intended cue recipient and retention interval, Experiment 1. Error bars represent standard error of the mean.

Table 2. Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 1 as a function of test number (1 vs. 2) and intended recipient of cues (self vs. other).

Fixed effect	$\hat{\beta}$	*SE*	Wald *z*	*p*
Intercept (baseline recall)	−1.325	0.155	−8.56	<.001
Test number	−1.181	0.080	−14.68	<.001
Intended recipient of cues	−0.034	0.078	−0.43	.664
Test number × Intended recipient	0.106	0.157	0.67	.501

Note: Total number of observations was 4480, log-likelihood = −2250.

in Table 3, recall decreased from the first test to the second, the three cue characteristics predicted recall, and none of the cue characteristics interacted with retention interval. More specifically, recall was greater when (a) cue-to-target associative strength was higher, (b) number of associates was lower, and (c) cue commonality was higher. The cue characteristics were not involved in any statistically significant interactions, meaning that the impact of each cue characteristic did not depend retention interval.

Discussion

In Experiment 1, participants generated memory cues for themselves or for others and we tested the effectiveness of the two kinds of cues at medium and long retention intervals. Participants generated different cues for themselves than for others. Cues intended for others showed greater cue-to-target associative strength, were less distinctive, and were less idiosyncratic (more common) than cues intended for oneself. The intended recipient of the cue, however, did not affect the efficacy of the cues. Cues for self and cues for others did not yield different recall at medium or long lags. The intended recipient of the cue likely did not affect recall because effective cue characteristics did not change with retention interval. Cue-to-target associative strength, distinctiveness, and cue commonality were all significantly associated with the effectiveness of cues, but the importance of the cue characteristics did not change with retention interval. In other words, cue characteristics that were beneficial at a 3 d retention interval were similarly beneficial at the 1 year delay.

The intended recipient also did not impact the efficacy of cues because cues for self and for others traded off between helpful characteristics. More specifically, cue-to-target associative strength supported recall, but so did distinctiveness. When switching from generating cues for themselves to generating cues for others, participants reduced the distinctiveness of cues but increased the cue-to-target associative strength. The intended recipient of cues did not change the efficacy of the cues because participants switched between two effective cue characteristics. Before interpreting these data further, we replicated this experiment with a different population in Experiment 2.

Table 3. Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 1, including cue characteristics.

Fixed effect	$\hat{\beta}$	*SE*	Wald *z*	*p*
Intercept (baseline recall)	−1.385	0.140	−9.88	<.001
Test number	−1.302	0.089	−14.65	<.001
Cue-to-target associative strength	5.32	0.362	14.70	<.001
Number of associates	−0.083	0.007	−12.14	<.001
Cue commonality	0.657	0.323	2.03	0.042
Test number × Cue-to-target assoc. str.	−0.399	0.645	−0.62	0.53
Test number × Num. assoc.	0.0001	0.013	0.006	0.995
Test number × Cue common.	0.400	0.545	0.73	0.46

Note: Total number of observations was 4480, log-likelihood = −2038.

Experiment 2

In Experiment 2, we attempted to replicate the results of Experiment 1. We recruited participants through Amazon Mechanical Turk, which may allow us to test a broader population (Difallah et al., 2018) and potentially increase the participant completion rate across the three sessions.

Method

Participants. We recruited 55 participants to complete the cue generation phase, 29 identified as female, 22 as male, and 4 as not reported. The average age of the sample was 35.57 ($SD = 9.54$). Of the 55 participants who completed the first phase, 48 completed the first test at the retention interval of 2 days. Thirty-six participants completed the final test at the year retention interval.

Design. The experimental design was identical to Experiment 1.

Materials. We narrowed down the list of target items to include 80 items from the list used in Experiment 1.

Procedure. The procedure was similar to Experiment 1 except that participants were paid for completing each phase of the experiment ($2 for cue generation, $2 for the first test, and $5 for the test after the year). Participants were reminded up to 2 times to complete each test portion of the experiment.

Results

Cues for self and others. We first examined whether participants differentiated between the cues intended for themselves versus others. Data are shown in Table 1 and replicate those from Experiment 1. Participants generated different kinds of cues for self and for others, as indicated by significant differences on all cue characteristic measures. Participants created cues that with fewer associates, weaker cue-to-target associative strength, and greater uniqueness for themselves than for others.

Cued Recall. As in Experiment 1, we next examined how effectively cues supported recall of target words as a function of test number (1 vs. 2) and intended recipient of cues (self vs. other). Results are shown in Figure 3. As in Experiment 1, we ran a linear mixed effect model to analyse the data. The results are shown in Table 4. Only

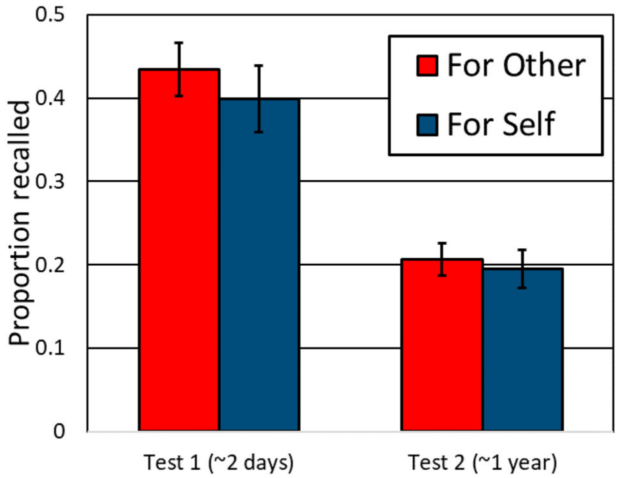

Figure 3. Mean proportion recall as a function of intended cue recipient and retention interval, Experiment 2. Error bars represent the one standard error of the mean above and below the sample mean.

the test number significantly impacted recall, such that participants remembered fewer items at the year-long retention interval than the 2-day retention interval. Neither intended recipient nor the interaction between test number and intended recipient reached significance.

Relation of cue characteristics to recall. Finally, we examined what cue characteristics related to recall of the targets and whether those cue characteristics changed with retention interval. As in Experiment 1, we conducted a linear mixed model predicting recall from test number (1 or 2), the three cue characteristics, and the interactions between the cue characteristics and test number. The results, shown in Table 5, reveal that recall was greater with greater cue-to-target associative strength and with fewer numbers of associates. Further, the impact of number of associates (i.e., distinctiveness) grew across retention interval.

Table 4. Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 2 as a function of test number (1 vs. 2) and intended recipient of cues (self vs. other).

Fixed effect	$\hat{\beta}$	*SE*	Wald *z*	*p*
Intercept (baseline recall)	−1.028	0.159	−6.48	<.001
Test number	−1.364	0.070	−19.36	<.001
Intended recipient of cues	−0.124	0.068	−1.81	.070
Test number × Intended recipient	0.105	0.136	0.77	.442

Note: Total number of observations was 5600, log-likelihood = −2918.

Table 5. Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 2, including cue characteristics

Fixed effect	$\hat{\beta}$	*SE*	Wald *z*	*p*
Intercept (baseline recall)	−1.105	0.145	−7.61	<.001
Test number	−1.506	0.079	−19.14	<.001
Cue-to-target associative strength	5.093	0.314	16.23	<.001
Number of associates	−0.100	0.006	−16.38	<.001
Cue commonality	0.069	0.301	0.23	0.82
Test number × Cue-to-target assoc. str.	−0.468	0.548	−0.85	0.39
Test number × Num. assoc.	−0.023	0.011	−2.04	0.04
Test number × Cue common.	0.283	0.518	0.55	0.58

Note: Total number of observations was 5600, log-likelihood = −2630.

Discussion

Experiment 2 largely replicated the findings of Experiment 1. Participants distinguished between cues for self and for others by increasing the cue-to-target association, decreasing the distinctiveness, and reducing the idiosyncrasies of cues for others. Despite the differences in cues generated, the intended recipient of the cue did not affect recall at medium or long retention intervals. Further, cue-to-target associative strength and number of associates were both related to a cue's effectiveness. In contrast to Experiment 1, the impact of number of associates on recall grew across retention intervals: At the second test, the distinctiveness of a cue was even more important than during the first test. In contrast to predictions, however, the results from Experiment 2 suggest that distinctiveness may matter more at longer retention intervals, which would suggest that cues for self could become especially effective at extended delays. However, the interaction between retention interval and number of associates did not reach significance in Experiment 1 (and was small in Experiment 2), so we hesitate to over-emphasize this inconsistent result.

Even though Experiment 2 hints that the effectiveness of cues may somewhat depend upon retention interval, we found no evidence that recall differs based upon the kinds of cues that learners generated or that recall depended upon the interaction of cue type with retention interval. As in Experiment 1, learners traded off distinctiveness for cue-to-target associative strength when generating cues for others. In other words, cues for self and cues for others were effective for different reasons. Both distinctiveness and cue-to-target associative strength support recall at short and long retention intervals. Effective cue characteristics did not change enough with retention interval to affect the efficacy of cue condition.

Some prior research has examined the impact of cues for self and others at very short retention intervals (i.e., within the same experimental hour as cue generation: Tullis & Benjamin, 2015b) by asking participants to generate one cue for themselves and one cue for others. Participants generated the same cues for themselves as for others for about half of the target items. However, *when learners produced different cues for self and others,* cues generated for others were less effective for one's own memory than cues generated for oneself. Our experiments differ from this prior research because we only elicited one cue per target, so we cannot conditionalize upon when participants distinguished between cues for self and others. Further, our first test happened 3 days after the cues were generated, on average. It is possible that we did not measure the efficacy of self-generated cues soon enough after they were generated to find differences between cues for self and for others.

In the third experiment, we included two additional conditions that have been examined in prior research and may elicit cues with greater cue-to-target associative strength than those from the prior two experiments. Further, we explicitly tested how stable cues are over time by measuring how frequently participants generated the same cue for the same target across a month retention interval.

Experiment 3

In Experiment 3, we directly tested the stability of cues over time by requiring participants to generate cues twice for the same targets over time. If cues generated for others rely upon stable semantic knowledge, participants should generate the same cue for a target across long lags. We additionally included conditions in which participants generated *descriptions* of the targets rather than mnemonic cues for the targets. Comparing the effectiveness of "descriptions" to "memory cues" assesses whether participants distinguish between characteristics that support memory compared to characteristics that describe a target. If participants have accurate metacognition about their memories, mnemonic cues should support cued recall better than descriptions. Further, prior research indicates that one-word *descriptions* are less distinct than *mnemonic cues*, but have similar cue-target associative strength (e.g., Tullis & Benjamin, 2015a). Experiment 3, then, allowed us to test the impact of distinctiveness on long-term retention.

In addition to the description condition, we also included a *focused description* condition, in which participants were instructed to generate a descriptive word that they would be likely to generate again later. Prior research suggests that under these focused description instructions, participants output descriptions that are more stable in time than descriptions of the target items (Mäntylä & Nilsson, 1988). If participants have accurate insight into what kinds of knowledge are stable in their minds over time, they will generate the same description for the same target over long lags more frequently in the focused description condition than the broader description condition. Manipulating the stability of cues through different instructions to participants may allow us to test whether stability of cues contributes to their effectiveness. Focused descriptions have been shown to be more helpful for cued recall than broader descriptions in prior research. For example, when provided with self-generated focused descriptions, memory performance only dropped from 95% to 80% over six weeks; when provided with self-generated spontaneous descriptions, performance dropped from 80% to 40% over the same time (Mäntylä & Nilsson, 1988).

We are interested in three specific comparisons among our conditions. First, to replicate and extend the prior two experiments, do cues for self and cues for others differ in characteristics, stability, and mnemonic effectiveness? Second, to replicate prior research (e.g., Mäntylä & Nilsson, 1988), do descriptions and focused descriptions differ in characteristics, stability, and mnemonic effectiveness? In other words, can learners assess what kinds of descriptions will be stable in time and does this stability contribute to effective recall? The distinction between descriptions and focused descriptions may reveal how effectively participants can take the perspective of their future selves. Prior research about the stability bias in metacognition suggests that learners have trouble anticipating changes in the cognition, as they make judgments based upon their immediate subjective experience (cognitive egocentrism) rather than on beliefs about future events (Kornell & Bjork, 2007; Kornell et al., 2011), but Mäntylä and Nilsson (1988) suggests that participants can decipher between focused and unfocused cues. Finally, in our third comparison, we examine whether cues for self and descriptions differ in characteristics, stability, and mnemonic effectiveness. Prior research has shown that self-generated mnemonic cues support recall more than descriptions at short retention intervals (Tullis & Benjamin, 2015a).

Method

Participants. We recruited 409 participants from Amazon Mechanical Turk to participate. Participants were paid $2 to complete the first day of cue generation and were given a bonus of $2 to complete the second portion of the experiment at a retention interval of 3 weeks. Participants were contacted twice to complete the second day of the experiment and 315 participants completed the second session.

Materials. The study list comprised 150 nouns from the University of South Florida Free Association Norms (Nelson et al., 1998). The nouns included a wide variety of topics.

Procedure. Participants completed the experiment online. Participants were randomly assigned to four conditions: cue generation for self, cue generation for others, description generation, and focused description generation. The cue generation conditions mimicked those from the prior two experiments. In the description condition, participants were told to generate one word that describes the target word (as in prior research: Tullis & Benjamin, 2015b). Specifically, participants received the following instructions in the description condition:

> For each target word, you will generate some aspect of the word that constitutes an appropriate description of the target item. This aspect or description can be created according to your own life experiences … For each target in the list, we ask that you type in one word that, according to your own experiences, describes the target word or is an aspect of the target word. You can use any description. However, the target word cannot serve as its own description.

Finally, in the focused description condition, participants were asked to generate descriptions of the target items that they would likely generate again under different circumstances (as in prior research: Mäntylä & Nilsson, 1988). Participants in the focused description condition received specific instructions to "Please generate 'FOCUSED' descriptions of the targets: descriptions you would likely generate again in a different situation … Remember, generate FOCUSED descriptions of the target items (descriptions that you would likely generate again under different circumstances)." All participants were instructed to generate a single English word that was not the target word for each target item and were told to expect a memory test. Participants generated cues or descriptions for 100 items during the first experimental session.

Three weeks later, participants were contacted to complete the follow-up test. During this second phase, participants were assigned to the same condition as during the initial phase. They were given 100 target words (50 novel targets and 50 that appeared during the first experimental session). As during the initial phase, they generated a single word to cue or describe the target. Participants were not told that half of the targets were items that they had seen on the first day, so there were no instructions about how their cues generated on the second day should relate to the cues generated on the first day. Finally, participants' memory was tested for the 50 targets that they saw during the initial phase but did not appear during the generation phase of day 2. During this test, the word generated during the first day was presented to learners and they had to recall the corresponding target. Targets from the first day were either assigned to the memory test or to the cue generation procedure during the second day; no target went through both phases on the second day.

Results

Cue characteristics. To view the entire sample of cues generated across description and cue conditions, see the data uploaded to the Open Science Framework. Examples of descriptions for targets include "cool" for "shade", "yellow" for "onion", and "shiny" for "jewelry". Examples of focused descriptions for targets include "white" for "snow" and "squarish" for "rectangle". We examined whether participants differentiated between the cues generated across conditions and the data are shown in Table 6. A one-way ANOVA on cue-to-target associative strength from cues generated during the first session showed a main effect of condition, $F(3, 405) = 25.25$, $p < .001$, $\eta^2 = .16$. Specific follow-up tests showed that cues for others had greater cue-to-target associative strength than cues for self, $t(205) = 3.41$, $p < .001$, $d = 0.47$. Cue-to-target associative strength did not significantly differ between descriptions and focused descriptions, $t(200) = .04$, $p = .97$, $d = 0.008$. Cues for self showed greater cue-to-target associative strength than descriptions, $t(215) = 3.87$, $p < .001$, $d = 0.57$.

A one-way ANOVA on the number of associates revealed a significant effect of condition, $F(3, 405) = 8.20$, $p < .001$, $\eta^2 = .06$. Cues for others did not have significantly greater number of associates than cues for self, $t(205) = 1.27$, $p = .20$, $d = 0.18$. Further, descriptions and focused descriptions did not differ in the number of targets associated from the cue, $t(200) = .20$, $p = .84$, $d = 0.03$. Descriptions had a greater number of potential associates than cues for self, $t(215) = 2.65$, $p = .009$, $d = .55$.

Finally, a one-way ANOVA on cue commonality indicated a significant effect of condition, $F(3, 405) = 7.24$, $p < .001$, $\eta^2 = .05$. Cues intended for others were more common than cues intended for oneself, $t(205) = 3.98$, $p < .001$, $d = 0.59$. Cue commonality did not significantly differ between descriptions and focused descriptions, $t(200) = 1.13$, $p = .26$, $d = 0.16$. Finally, cues for self were less common than descriptions, $t(215) = 3.17$, $p = .002$, $d = 0.30$.

Stability of cues. Next, we examined how frequently participants supplied the same cue to the same target across the two generation phases which were separated by 3 weeks. The results, shown in the left panel of Figure 4, reveal a significant effect of condition on the probability of supplying the same cue across lag, $F(3, 308) = 3.05$, $p = .03$, $\eta^2 = .03$. Planned comparisons indicated the cues generated for others were more stable over time than cues generated for the self, $t(161) = 2.67$, $p = .008$, $d = .42$. Focused descriptions did not differ in stability from descriptions, $t(147) = .49$, $p = .63$, $d = .08$. Finally, descriptions did not significantly differ from cues for the self, $t(158) = .54$, $p = .59$, $d = 0.09$.

Cued recall. Finally, we examined how effectively the different conditions supported cued recall on the memory test at the end of the second experimental session. Condition had a significant impact on how much learners remembered, $F(3, 308) = 14.49$, $p < .001$, $\eta_p^2 = .12$. Recall did not significantly differ between cues intended for others compared to cues for the self, $t(161) = .55$, $p = .59$, $d = .06$. Focused descriptions did not support recall better than descriptions, $t(147) = .60$, $p = .55$, $d = .09$. Cues for the self supported recall better than descriptions, $t(158) = 4.66$, $p < .001$, $d = .73$.

Discussion

In Experiment 3, we directly tested the stability of mnemonic cues over time and whether the stability of cues determines their eventual effectiveness. The instructions shaped the characteristics of the cues that participants generated. For

Table 6. Means and standard deviations in Experiment 3 of cue characteristics for cue-target Pairs across the four conditions.

Cue characteristic	Cue For Others	Cue For Self	Description	Focused Des.
Cue-to-target associative str	.09 (.03)	.07 (.03)	.06 (.02)	.06 (.03)
Number of associates	9.35 (2.47)	8.91 (2.49)	10.30 (2.53)	10.23 (2.17)
Cue commonality	.12 (.03)	.10 (.04)	.11 (.03)	.10 (.03)

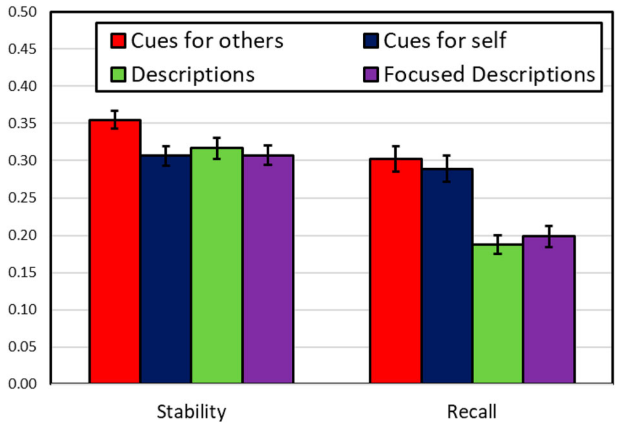

Figure 4. Mean proportion stability (i.e., proportion of the same cues generated) and Mean Proportion Cued Recall as a Function of Condition, Experiment 3. Error bars represent the one standard error of the mean above and below the sample mean.

example, as in prior experiments, cues for others showed greater cue-to-target associative strength and greater commonality than cues for oneself. Further, cues for self showed stronger cue-to-target associative strength and were more distinct than descriptions. We found no differences in descriptions and focused descriptions.

The different kinds of cues generated yielded differential stability over time. The most stable cues generated were cues that were intended for others. When participants generated cues for others, the data suggest that they tapped into relatively stable knowledge structures to do so. The results imply that people have a relatively stable conception of others' mindsets and what cues would be useful for others. In contrast, cues for oneself (and descriptions of the targets) shifted significantly more than cues for others across the month delay. One's own mindset, then, seems to shift more than one's conceptions of others' mindsets. Again, in contrast to prior research, we find no difference in the stability of descriptions compared to focused descriptions (cf. Mäntylä & Nilsson, 1988).

The stability of the generated mnemonic cues did not directly cause differences in mnemonic effectiveness. The patterns for the stability of the cues differed from the pattern for the recall of the targets across conditions. While cues for others were the most stable over time, they were equally as supportive of recall as cues for the self. Further, both kinds of cues supported recall better than both kinds of descriptions, even though both kinds of descriptions were as stable as cues for the self. These results suggest that the bottleneck of cues is not their stability, but potentially their ability to trigger recall of the appropriate target word (i.e., their distinctiveness). Prior research suggests that mnemonic cues must be both memorable and decipherable (Dunlosky et al., 2005); stability may reflect the memorable component of cues. Distinctiveness may most closely relate to the decipherability of those cues.

General diascussion

Across three experiments, we tested the effectiveness of mnemonic cues across long retention intervals. First, participants consistently generated different cues for oneself than for others. Participants generated cues for others that had greater cue-to-target associative strength, less idiosyncrasies, and greater stability than cues for oneself. Second, the intended recipient of the cues did not change the efficacy of cues at medium or long retention intervals. Prior research shows that the intended recipient of the cue matters more when cues are given to others than when they are given back to the generator (Tullis & Benjamin, 2015b). More specifically, when cues are given to others, cues intended for others are more effective than cues intended for oneself, but when cues are given back to the generator, the intended recipient does not significantly impact cued recall. In the current experiments, the generator always received their own cues during the test, which reduces the impact of the intended recipient of the cue. Further, the equivalent effectiveness of these two kinds of cues likely reflects the trade-off between effective cue characteristics. Cue-to-target associative strength and distinctiveness both support recall. Cues for others showed greater cue-to-target associative strength, but less distinctiveness than cues for the self. Finally, the differential effectiveness between cues for self and cues for others may be too small to detect within our cued recall procedure, which uses only single words for both cues and targets.

Generated cues supported recall much more than descriptions, which indicates that participants could somewhat effectively anticipate their future recall needs during mnemonic cue generation. Participants understood that mnemonic cues are more effective when they are can distinguish between potential targets (i.e., when they are distinctive). In contrast to predictions, effective cue characteristics did not change with retention interval. Cues that were effective at medium retention intervals retained their effectiveness over a year. These results suggest that anticipating future mindsets or taking one's future perspective are not the limiting factors when creating effective mnemonic cues for simple materials. The case may be different for more complex materials in everyday life, such as journals, class notes, or comments in computer code. In an informal pilot survey of undergraduates and their friends and family, only 29% of 137 said they had never struggled to figure out a note or reminder they had previously written for themselves, indicating that this is a widespread experience.

Stability of perspectives

Cue condition affected the stability of the cues that participants generated, but stability did not drive the effectiveness of a cue. Participants' cues for others were more stable than cues for themselves, which indicates that participants access a mental model of others' knowledge that is more stable over time than one's own mental state. Conceptions of others change less quickly than one's own idiosyncratic mental state does. Participants tap into normative knowledge about the targets and normative knowledge does not shift as quickly as one's current mental state. This difference may be an interesting avenue to explore in the future; are predictions about others' memories less affected by current and idiosyncratic processing? Taking perspective of a different learner may allow one to dissociate one's immediate ongoing experiences when predicting future performance.

Instructing participants to generate "focused" descriptions of the target items did not affect description generation in any of our measures. While we affected the stability and kinds of cues generated by changing the intended recipient of mnemonic cues, the kind of description requested did not affect description generation. Participants were not able to tap into their own persistent mental states and knowledge when instructed to do so. This suggests that participants may struggle to anticipate their own future mental states and may not be able to predict how or in what ways their mental states will shift, which echoes the stability bias in metacognition (Kornell & Bjork, 2009). Interestingly, by requesting that participants take another's perspective, participants' cues become significantly more stable over time.

Control of the retrieval environment

Cue generation addresses an important gap in our understanding of metacognition: how learners control their *retrieval environments*. Learners control their study and encoding by, for instance, choosing study strategies or allocating the amount of time spent on different material (see Finley et al., 2009 for a review). However, learners may also regulate their later retrieval, such as by controlling what cues will be available to them (as in the present study) or by adjusting their retrieval strategies (e.g., Finley, 2012; Fraundorf & Benjamin, 2016). Growing evidence suggests that, even with shifting mental contexts (Kornell & Bjork, 2009), generators can select cues that are more effective for themselves than others can (Tullis & Fraundorf, 2017). We additionally showed that self-generated cues more effectively cue memory than self-generated descriptions at long retention intervals. Our data provide more evidence of the importance of allowing learners to exercise control over their own memories and learning (Tullis & Benjamin, 2011).

We have argued that effectively anticipating and controlling available retrieval routes (i.e., retrieval environments) during encoding, however, may be an especially challenging form of self-regulated learning because learners must take perspective of their selves in the future. Yet, our current data may indicate that taking the perspective of oneself in the future is not the limiting factor in producing effective mnemonic cues. The stability of cues, which may indicate an aspect of the ability to take one's future perspective, does not determine recall. Further, even when cue-to-target associative strength is increased in the cues for others conditions, this does not significantly increase recall of corresponding targets. Producing effective cues that reinstate encoding conditions may be more important than cues that are stable over time.

The real-life experience of failing to decipher one's old notes presents an interesting phenomenon worthy of laboratory study, but one that may be difficult to replicate using simple material in the lab. Using simple one-word targets and one-word cues may not provide the opportunity for learners to create cues that are idiosyncratic *enough* to work well in the short-term but more poorly in the long-term. The phenomenon may also be difficult to pinpoint in part because some idiosyncratic knowledge *does* in fact persist over very long periods of time (cf. permastore, Bahrick, 1984). Indeed, "secret questions" used for online security specifically attempt to leverage this kind of information (e.g., name of your first pet), in the hopes that it will still be retrievable even years later if one's password is forgotten. But even these are susceptible to reduced effectiveness as people age and change. For example, your favourite movie may well change over time, or the city in which you met your partner may change with new relationships.

Another important limitation of these experiments is that we provide participants with their cues during the memory test. Participants only need to decode or interpret the cue presented to them (Dunlosky et al., 2005); participants do not need to recall the cue. This mimics some real-world scenarios, like when people try to understand their written to-do lists or when people try to remember what a specific computer file name references. However, in many other scenarios, learners have to recall the cue before they can recall the target. For example, when trying to use mnemonics to remember chemistry facts, students would be given a chemistry question, they would need to recall the mnemonic cue they generated, and finally, they would need to decode that cue to recall the target information. Anticipating one's future mental states may play an important role in *recalling the cue*, even though we show little influence of that cue stability on target recall. With these data, we cannot assess how cue generation impacts learners' ability to retrieve the cue, which has sometimes been shown to limit how much learners can recall (Dunlosky et al., 2005). Understanding the processes, strengths, and weaknesses of cue generation may ultimately allow us to develop targeted interventions to help people generate better cues so that they can more effectively remember and apply information.

Note

1. Cues for others appeared in the South Florida Free Association Norms more frequently ($M = .73$, $SD = .08$) than cues for self (M

= .69, *SD* = .11). Conditionalizing analyses on just the cues contained within the database yields the same pattern of results in cue-to-target associative strength between conditions: Cues for others showed greater cue-to-target associative strength (M = .094, SD = .035) than cues for self (M = .081, SD = .031), $t(68) = 3.20$, $p = .002$, $d = 0.39$. Similar analyses for all three Experiments are available in supplemental materials.

Acknowledgements

Thanks to Scott Fraundorf for statistical advice.

Disclosure statement

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

Funding

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

Data availability statement

The data and materials for all experiments are (Tullis & Finley, 2021, February 4) available at https://osf.io/z4f6c/?view_only=a0fc8a6f83e149cca31cee2b3e855274

ORCID

Jonathan G. Tullis http://orcid.org/0000-0001-5237-1776

References

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(2), 126–113. doi:10.1037/0278-7393.1.2.126

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, *113*(1), 1–29. https://doi-org.libproxy.siue.edu/10.1037/0096-3445.113.1.1

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bellezza, F. S., & Poplawsky, A. J. (1974). The function of one-word mediators in the recall of word pairs. *Memory & Cognition*, *2*(3), 447–452. doi:10.3758/BF03196903

Benjamin, A. S. (2008). Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 175–223). Academic Press.

Benjamin, A. S., & Bird, R. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory & Language*, *55*(1), 126–137. doi:10.1016/j.jml.2006.02.003

Bloom, C. M., & Lamkin, D. M. (2006). The Olympian struggle to remember the cranial nerves: Mnemonics and student success. *Teaching of Psychology*, *33*(2), 128–129.

Difallah, D., Filatova, E., & Ipeirotis, P. (2018, February 5–9). Demographics and dynamics of Mechanical Turk workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 135–143). ACM.

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*(2), 389–400. doi:10.1037/0012-1649.41.2.389

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*(5), 369–377. doi:10.1037/h0046888

Finley, J. R. (2012). *Retrieval cue variability: When and why are two meanings better than one?* [Doctoral dissertation, University of Illinois at Urbana-Champaign]. Illinois Digital Environment for Access to Learning and Scholarship. http://hdl.handle.net/2142/31074

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm.. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 632–652. doi:10.1037/a0026215

Finley, J. R., & Benjamin, A. S. (2019). *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* (pp. 109–131). Springer Science & Business Media.

Fraundorf, S. H., & Benjamin, A. S. (2016). Conflict and metacognitive control: The mismatch-monitoring hypothesis of how others' knowledge states affect recall. *Memory*, *24*(8), 1108–1122. doi:10.1080/09658211.2015.1069853

Harris, J. E. (1980). Memory aids people use: Two interview studies. *Memory & Cognition*, *8*(1), 31–38. doi:10.3758/BF03197549

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. doi:10.3758/BF03200882

Jamieson, D. G., & Schimpf, M. G. (1980). Self-generated images are more effective mnemonics. *Journal of Mental Imagery*, *4*, 25–33.

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*(2), 187–194. doi:10.1037/0278-7393.31.2.187

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self–regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219–224. doi:10.3758/BF03194055

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, *138*(4), 449–468. doi:10.1037/a0017350

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*(3), 609–622. doi:10.1037/0278-7393.32.3.609

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. doi:10.1177/0956797611407929

Kuo, M., & Hooper, S. (2004). The effects of visual and verbal coding mnemonics on learning Chinese characters in computer-based instruction. *Educational Technology Research and Development*, *52*(3), 23–34. doi:10.1007/BF02504673

Levin, M. E., & Levin, J. R. (1990). Scientific mnemonomies: Methods for maximizing more than memory. *American Education Research Journal*, *27*(2), 301–321. doi:10.3102/00028312027002301

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 756–766. doi:10.1037/0278-7393.10.4.756

Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 66–71. doi:10.1037/0278-7393.12.1.66

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. doi:10.1111/j.1467-9450.1983.tb00504.x

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), 502–509. doi:10.1037/0278-7393.14.3.502

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5-6), 389–395. doi:10.1080/09658210244000216

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Quoidbach, J., Gilbert, D. T., & Wilson, T. D. (2013). The end of history illusion. *Science*, *339*(6115), 96–98. doi:10.1126/science.1229294

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. *Psychology of Learning and Motivation*, *14*, 207–262. doi:10.1016/S0079-7421(08)60162-0

Richmond, A. S., Carney, R. N., & Levin, J. R. (2011). Got neurons? Teaching neuroscience mnemonically promotes retention and higher-order thinking. *Psychology Learning & Teaching*, *10*(1), 40–45. doi:10.2304/plat.2011.10.1.40

Ryskin, R. A., Benjamin, A. S., Tullis, J. G., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General*, *144*(5), 898–915. doi:10.1037/xge0000093

Saber, J. L., & Johnson, R. D. (2008). Don't throw out the baby with the bathwater: Verbal repetition, mnemonics, and active learning. *Journal of Marketing Education*, *30*(3), 207–216. doi:10.1177/0273475308324630

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394. doi:10.1037/0033-2909.121.3.371

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, *64*(2), 109–118. doi:10.1016/j.jml.2010.11.002

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review*, *19*(4), 743–749. doi:10.3758/s13423-012-0266-2

Tullis, J. G., & Benjamin, A. S. (2015a). Cue generation: How learners flexibly support future retrieval. *Memory & Cognition*, *43*(6), 922–938. doi:10.3758/s13421-015-0517-3

Tullis, J. G., & Benjamin, A. S. (2015b). Cuing others' memories. *Memory & Cognition*, *43*(4), 634–646. doi:10.3758/s13421-014-0478-y

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), 540–552. doi:10.1037/xlm0000473

Tullis, J. G., & Finley, J. (2018). Self-generated memory cues: Effective tools for learning, training, and remembering. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 179–186. doi:10.1177/2372732218788092

Tullis, J. G., & Finley, J. R. (2021, February 4). *What characteristics make self-generated memory cues effective over time?* http://www.osf.io/z4f6c

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. doi:10.1016/j.jml.2017.03.003

Tullis, J. G., & Maddox, G. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition & Learning*, *15*(2), 129–154. doi:10.1007/s11409-020-09223-x

Tullis, J. G., & Qui, J. (2021). Generating mnemonics boosts recall of chemistry information. *Journal of Experimental Psychology: Applied*. Advance online publication. https://doi.org/10.1037/xap0000350

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, *71*(5), 673–679. doi:10.1037/h0023107

Van Etten, S., Freebern, G., & Pressley, M. (1997). College students' beliefs about exam preparation. *Contemporary Educational Psychology*, *22*(2), 192–212. doi:10.1006/ceps.1997.0933

Zhang, D., & Tullis, J. G. (2021). Personal reminders: Self-generated reminders boost memory more than normatively related ones. *Memory & Cognition*, *49*, 645–659. doi:10.3758/s13421-020-01120-7.