



Selecting effectively contributes to the mnemonic benefits of self-generated cues

Jonathan G. Tullis¹ · Scott H. Fraundorf²

Accepted: 21 September 2021
© The Psychonomic Society, Inc. 2021

Abstract

Self-generated memory cues support recall of target information more robustly than memory cues generated by others. Across two experiments, we tested whether the benefit of self-generated cues in part reflects a meta-mnemonic effect rather than a pure generation effect. In other words, can learners select better memory cues for themselves than others can? Participants generated as many possible memory cues for each to-be-remembered target as they could and then selected the cue they thought would be most effective. Self-selected memory cues elicited better cued recall than cues the generator did not select and cues selected by observers. Critically, this effect cannot be attributed to the process of generating a cue itself because all of the cues were self-generated. Further analysis indicated that differences in cue selection arise because generators and observers valued different cue characteristics; specifically, observers valued the commonality of the cue more than the generators, while generators valued the distinctiveness of a cue more than observers. Together, results suggest that self-generated cues are effective at supporting memory, in part, because learners *select* cues that are tailored to their specific memory needs.

Keywords Metacognition · Metacognitive control · Cue generation · Generation effect · Perspective taking

Introduction

Learners make many metacognitive choices while studying, and the choices that they make dictate how much they learn (Finley et al., 2009; Thiede et al., 2003). Broadly, learners select strategies during encoding that benefit later recall, including selecting which items to restudy (Kornell & Metcalfe, 2006; Tullis & Benjamin, 2012), allocating study time (Tullis & Benjamin, 2011), and choosing retrieval practice over restudy (Tullis et al., 2018, but see Karpicke, 2009). 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). While ample research shows that learners broadly make effective study choices about *encoding*,

less research examines how effectively learners choose or create their *retrieval* environments (but see Finley et al., 2009).

Although less research has examined control of retrieval, creating cues to support one's memory – one way that learners control their future retrieval environment – is nevertheless 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 colors 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). For instance, in a classic laboratory experiment, participants remembered 91% of words when they were prompted with self-generated cues, but only 55% of targets when prompted by cues generated

✉ Jonathan G. Tullis
jonathantullis@gmail.com

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

² Learning Research and Development Center and Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

by others (Mäntylä, 1986). 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). The benefits of self-generation persist over long retention intervals of up to 3 weeks after study (e.g., Bloom & Lamkin, 2006; Mäntylä, 1986; Mäntylä & Nilsson, 1988; but see Kibler & Blick, 1972). Further, self-generated cues have been shown to benefit memory for multiple types of materials, including simple words (Mäntylä, 1986; Mäntylä & Nilsson, 1983), foreign language vocabulary (Atkinson & Raugh, 1975), and complex science concepts (Levin & Levin, 1990; Richmond et al., 2011; Tullis & Qiu, *in press*). Here, we test competing ideas about *why* learners' self-generated cues can effectively support memory.

Why are self-generated cues effective?

One simple reason that learners' self-generated cues benefit memory may be that generating a cue, in and of itself, bolsters memory. Decades of research provides evidence for a *generation effect* in human memory: When learners generate all or part of a stimulus item, they remember that material better than material they only read (Jacoby, 1978; Hirshman & Bjork, 1988; Slamecka & Graf, 1978). Meta-analysis shows that generation benefits memory across different types of tests, to-be-remembered stimuli, learner populations, and retention intervals (Bertsch et al., 2007). Theories of the generation effect suggest that generation requires more cognitive work than reading, and the extra exertion of mental effort boosts memory (e.g., McFarland Jr. et al., 1980). As applied to learner-generated cues, learners may process to-be-remembered information more deeply when generating cues than when studying others' cues. Learners may make semantic, acoustic, or visual connections between the target information and their own knowledge during cue generation (Fisher & Craik, 1980). Some empirical data hint that generation contributes to the mnemonic benefits of cue generation: Learners spend more time generating their own memory cues than reading cues generated by others (Tullis & Benjamin, 2015a).

Beyond the well-established mnemonic effects of generation, self-generated cues may also confer a *metamnemonic* benefit. According to this *cue-selection hypothesis*, self-generated cues may be more effective at supporting memory than cues generated by others in part because learners can 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 select uniquely effective cues (Kuo

& Hooper, 2004; Symons & Johnson, 1997; Tullis & Finley, 2018). Prior research has not been able to untangle the differing contributions of generation and selection because participants only output a single cue for each target; in contrast, the current research requires participants to output multiple potential cues and select which self-generated cue would be most effective.

In the experiments presented here, we examined how – and how effectively – learners generate and select cues for themselves. Our primary interest was in testing three critical predictions of the *cue-selection* hypothesis. First, the cue-selection hypothesis implies that learner-selected cues should support memory better than unchosen cues, even when controlling for generation effects. We tested this prediction by comparing learners' memory when given the cues they generated and selected versus memory when given cues they generated but did not select. Honoring learners' choices from their own self-generated cues should yield better recall than dishonoring those choices. Second, the cue-selection hypothesis suggests that learners consider multiple self-generated cues before selecting an especially effective one. This hypothesis implies that learners choose effective cues rather than the first cues that come to mind. More specifically, learners should benefit from selecting their own cues even when we control for output order of those cues.

Finally, for cue selection (and privileged access to one's own idiosyncratic mental state) to account for the efficacy of self-generated cues, learners should be particularly adept at selecting cues for their *own* memory as compared to others'. We tested this by comparing learners' selections of cues for their own memory to observers' abilities to select among the same cues. We tested these three critical predictions about cue selection by requiring cue generators to output multiple cues per target (in contrast to all prior research (Tullis & Benjamin, 2015a, 2015b; Tullis & Fraundorf, 2017), which elicited only a single cue per target). To preview, our data support each of these predictions and reliably show that cue selection contributes to the effectiveness of learner-generated cues.

Cue characteristics

While our primary interest was in *whether* cue selection contributes to the benefit of learner-generated cues, we further examine *how* learners are able to select better cues for themselves than others. The differential effectiveness of self-selected and other-selected cues may arise because learners may value different cue characteristics when choosing cues for different people.

Past work (Tullis & Benjamin, 2015a, 2015b; Tullis & Fraundorf, 2017) has identified three characteristics of effective cues that generators may value when selecting effective cues: cue-to-target associative strength, idiosyncrasy, and

distinctiveness. First, cues with greater cue-to-target associative strength more effectively bring the target to mind than cues with lesser cue-to-target associative strength. Choosing cues that strongly evoke the target items may help learners easily decode their cue during retrieval (Dunlosky et al., 2005).

Second, distinctive cues may enable stronger recall than common cues because they may have fewer extra-list associations (Einstein & McDaniel, 1990) and can limit the range of potential targets. If a cue is distinct (i.e., it is not associated to many potential targets), it narrows the search field during retrieval and improves recall (Hunt & Smith, 1996). For example, *beatles* may serve as a good cue for the target *band* for me because *beatles* does not point to many potential targets. In contrast, *school* may be a poor cue for *band* because *school* points to many possible target words. Distinctiveness may be the single most important attribute in determining whether a cue effectively enables target recall or not (Nairne, 2002). Finally, when generating and selecting their own cues, learners may be able to utilize idiosyncratic cues that rely upon their own episodic experiences. Selecting cues that tie into one's own idiosyncratic past experiences may be mnemonically beneficial, as in the *self-reference effect* (e.g., Symons & Johnson, 1997), because cues that tie into one's own personal experiences engender organized and elaborate processing.

The cue-selection hypothesis posits that learners can select effective cues for themselves because they appropriately value these cue characteristics. More specifically, when choosing cues for themselves, learners may not place as much value on normative relations and commonality of cues because they can base their cues in their own experiences and idiosyncratic associations. However, when choosing cues for others, learners may select common cues for others that have strong normative associative relations to the targets to ensure some association exists for others. We test these ideas across the experiments by comparing characteristics of cues that learners choose for themselves with those that they choose for others.

In sum, across the experiments reported here, learners generated multiple mnemonic cues to help remember a series of target items. Learners then selected which of their generated cues would best support their memory. We compared the effectiveness of their selected cues with those picked randomly from their generated list and with those picked by a different learner. In support of the cue-selection hypothesis, learners picked more effective mnemonic cues than cues selected at random or by other learners.

Experiment 1a

In [Experiment 1a](#), we tested the hypothesis that participants would choose mnemonic cues that support their memories

from a list of self-generated options more effectively than random selections would. Participants studied a list of target words and they reported all the potential mnemonic cues that came to mind for each cue. Then, participants selected the cue that they believed would most effectively support their memory for each target word. We assessed the effectiveness of these cue selections by comparing performance on a subsequent cued recall between an honor and dishonor condition (for other research using honor/dishonor paradigms to test the efficacy of metacognitive control, see Kornell & Metcalfe, 2006; Kimball et al., 2012). In the honor condition, participants were given the cue they had thought would be most effective for supporting memory during the later cued-recall test. In the dishonor condition, participants were given a cue they had generated but did not select as most effective. If learners are adept at selecting cues that particularly support later recall – as predicted by our cue-selection hypothesis – then their memory performance should be greater when given cues that they selected than when given other, unchosen cues. Critically, *all* cues (even those presented in the dishonor condition) are self-generated, so any differences in cued recall across conditions must be driven by learners' selections from among the potential cues rather than the process of generation itself.

Method

Participants Our primary interest was the within-subjects comparison between the honor and dishonor conditions. Given a medium effect size (Cohen's $d = 0.5$), as suggested by prior studies (Tullis & Benjamin, 2015a), a power analysis suggested that we could obtain 80% power with 34 participants in a paired-samples t -test; our mixed-effects model should, if anything, have even higher power (Quené & van den Bergh, 2004). Thus, 34 undergraduate, introductory educational psychology students participated in order to earn partial course credit.

Materials Fifty words were collected from the University of South Florida Free Association Norms (Nelson et al., 1998). We chose to-be-remembered words that are likely relevant to a college student's life such that participants could potentially have personal experiences with them. Targets included words like "library," "hobby," and "professor"; a full list of materials and data are available on the Open Science Framework at https://osf.io/muhtc/?view_only=36d20866557444fb914c3e26784fafa0.

Procedure Participants completed the experiment on a desktop computer while up to three other participants participated on other computers in the same room. Participants were instructed to remember a list of target words for an upcoming memory test. Participants were told that they would generate a list of potential cues to help them remember each target. They

were instructed that they would get a cue back during the test and would have to recall the corresponding target word. Participants could generate any cue that was a single, English word that is not a form of the target (i.e., a plural or a misspelling of the target). Participants typed in every single cue that came to mind in the order that they came to mind. Participants were required to generate at least two possible cues for each target but were encouraged to provide as many as came to mind. As participants entered each potential cue, the list of generated cues was displayed on the bottom of the screen in the order that they output the cues. When they were finished typing all the possible cues that came to mind, participants were instructed to type "finished." Next, participants had to pick the cue that they believed would be the most helpful during the test by typing in that cue. The target stayed on the screen throughout the cue generation and selection process. After they generated and selected cues for each target, they completed an unrelated motor movement task for approximately 10 min.

Finally, participants took a cued-recall test on the target information. During the cued-recall test, a cue that they generated was displayed on the screen and participants tried to recall the corresponding target word. For a random half of the targets, participants received the cue they had selected. For the other half of the targets, participants' choices were dishonored; they received one of the cues they generated but did not select. The dishonored cue was chosen at random from all the non-selected cues the participant provided.

Results

Number of cues output First, we examined how many potential cues participants output. Participants supplied, on average, 3.81 ($SD = 1.36$) cues per target item; Fig. 1 depicts the distribution across participants of the average number of cues generated per target.

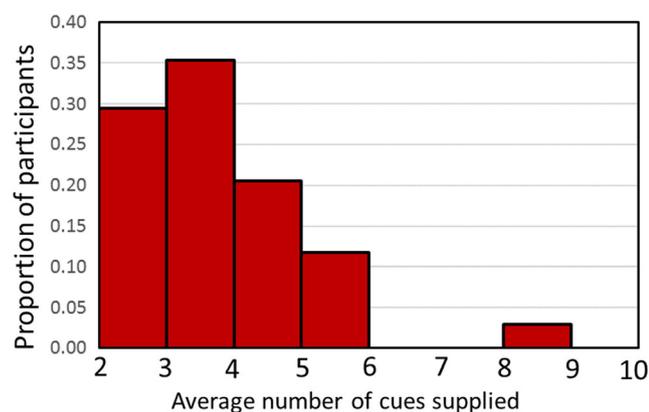


Fig. 1 Distribution of the average number of cues supplied per target in Experiment 1a

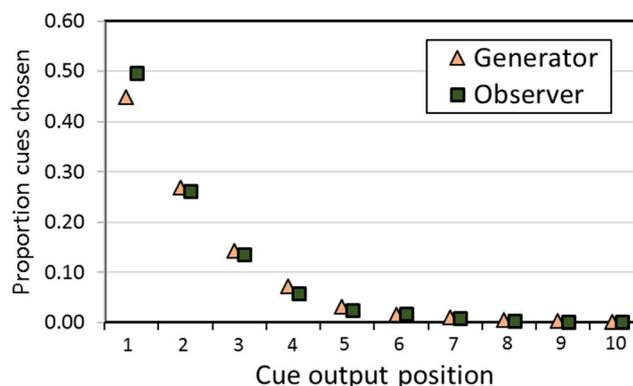


Fig. 2 Proportion of cues chosen by output position by the cue generator (Experiment 1a) and by an observer (Experiment 1b). Data points are jittered horizontally to display each clearly

Effectiveness of cue choice Next, we examined whether participants chose cues effectively by comparing cued recall between honored and dishonored conditions. To test this, we fit a mixed logit model with the accuracy of recall as the dependent variable and the condition as an independent variable. One explanation for any potential benefit of honoring learners' cue choices within this experiment is simply that requiring learners to output multiple cues prompts them to generate cues they would never consider as possible cues. Honoring learners' choices would then tend to favor earlier – and presumably better – cues. In the dishonor condition, by definition, cues were equally likely to come from any output order position and include ineffective cues output later than effective cues. In fact, as Fig. 2 illustrates, learners do select cues that were generated earlier at greater rates than those output later. To control for the possibility that output order is implicated in cue selection, we added output order and its interaction with condition as additional independent variables to our model.

Table 1 displays the results of this model. The model revealed a significant effect of condition: The odds of correct recall were 3.77 (95% CI: [2.84, 4.99]) times greater when a participant's cue choice was honored ($M = 0.82$) than when it was dishonored¹ ($M = 0.57$), $z = 9.26$, $p < 0.001$. In fact, only one participant showed better memory for dishonored cues than honored cues. Further, this effect cannot be attributed to output order; there was no significant main effect of output order, nor did it significantly interact with condition. Indeed, plotting the proportion correct cued recall by output position and honor condition (Fig. 3, below) shows that honored cues outperformed dishonored cues at every output position.

Relation of cue characteristics to recall Why did honoring participants' cue choices improve recall? Participants may have chosen cues with characteristics that effectively cued

¹ For all models reported here, the same conclusions were reached if output order was excluded from the model entirely.

Table 1 Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 1 as a function of condition

Fixed effect	β	SE	Wald z	p
Intercept (baseline log odds of recall)	1.019	0.136	3.62	< .001
Honor condition	1.326	0.143	9.27	< .001
Output order	0.016	0.053	0.31	.760
Honor condition \times Output order	0.003	0.091	0.03	.973

Note. SE = standard error

the target. To test this hypothesis, we first examined which cue characteristics were associated with a higher probability of recall (e.g., *diagnosticity*: Van Loon et al., 2014); then, we examined whether those same cue characteristics increased the probability that participants would *select* those cues (i.e., *utilization*; van Loon et al., 2014).

As in prior research analyzing cue generation (Tullis & Benjamin, 2015a; Tullis & Benjamin, 2015b; Tullis & Fraundorf, 2017), we examined three cue characteristics. *Normative cue-to-target associative strength* was operationalized as the proportion of a sample of participants who freely associate the target with the cue from the South Florida Free Association Norms (Nelson et al., 1998). *Cue distinctiveness* was determined based on the number of words associated (in the South Florida Free Association Norms) from each cue. Finally, we computed *cue commonality* as proportion of participants in this experiment (calculated across all output positions) who supplied the same cue for each target item. Smaller cue commonality indicates that fewer others reported the same cue and the cue is more idiosyncratic.

Because these cue characteristics are originally measured on different scales (e.g., the number of associated words for cue distinctiveness versus a proportion of participants for cue commonality), we z-scored them so that all of the regression coefficients correspond to the effect of a 1-standard-deviation

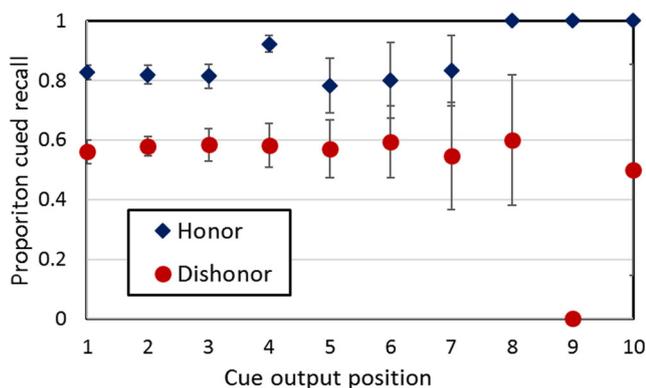


Fig. 3 Proportion of correct cued recall as a function of honor/dishonor condition and cue output position. Error bars show standard errors of the mean; error bars get progressively wider with output position due to the reduced number of cues supplied at those positions

(SD) change in the cue characteristics (see also Tullis & Fraundorf, 2017). Table 2 displays the mean² for these cue characteristics (in both the original units and z-scores) for items that the learner eventually remembered and for items that the learner eventually did not remember. However, the bivariate relationships must be interpreted with caution; because these cue characteristics are not orthogonally manipulated, it is possible that the apparent presence or absence of one effect taken alone (e.g., an effect of cue distinctiveness) could instead result from a partial confound with some other cue characteristic (e.g., cue commonality). Rather, the cue characteristics can be better understood in a multiple regression that tests the effect of one cue characteristic while holding others constant.

We thus used a mixed logit model, displayed in Table 3, to simultaneously examine the relationship of these three cue characteristics – as well as the control variable of output order – to recall. The model indicated that all three cue characteristics influenced recall (replicating Tullis & Fraundorf, 2017), even when controlling for each other and for output order. Consistent with Nairne (2002), the strongest predictor of cue effectiveness was cue distinctiveness: a 1-SD decrease in the number of words associated with the cue (i.e., a more distinctive cue) increased the odds of correct recall by 1.52 times (95% CI: [1.33, 1.73]). There were also effects of associative strength and cue commonality: A 1-SD increase in associative strength increased the odds of correct recall by 1.31 times (95% CI: [1.09, 1.56]), and a 1-SD increase in cue commonality by 1.45 times (95% CI: [1.23, 1.71]). The order in which a cue was originally output was not significantly associated with its ability to later cue recall.

Relation of cue characteristics to selection Did learners select certain cues because they had the desirable characteristics of being distinctive, common, and strongly associated with the targets? We next examined utilization: which cue characteristics predicted whether participants would select a cue as one they wanted to cue their later memory. Table 4 displays the average characteristics of cues selected for later use versus cues not selected for later use (regardless of whether or not that selection was ultimately honored, since participants had no idea at the time of choice that some selections would not be honored).

Table 5 displays the results of a mixed-effects model of the (log) odds that a cue would be selected for later use as a function of the cue characteristics. Participants’ selections were sensitive to all three of the cue characteristics that

² Because cues beyond the tenth serial position were extremely rare (only 19 such cues across all trials and all participants), we could not validly model the effects of output order beyond the tenth serial position since regression inferences are not valid for values of the predictor variables that appear sparsely or not at all. Thus, in this and all following analyses, we included only cues up to the tenth serial position, eliminating fewer than 1% of the data.

Table 2 Means and standard deviations in Experiment 1a of cue characteristics for cue-target pairs that were and that were not remembered, measured in original units (top half) or standardized scores (bottom half)

Cue characteristic	Items remembered		Items not remembered	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cue-to-target associative strength	0.04	0.11	0.02	0.08
Number of associates	8.56	7.55	10.56	7.86
Cue commonality	16%	20%	11%	17%
Output order	2.25	1.44	2.35	1.42
Cue characteristic (<i>z</i> -scored)				
Cue-to-target associative strength	0.07	1.06	-0.17	0.81
Number of associates	-0.08	0.98	0.18	1.02
Cue commonality	0.07	1.04	-0.17	0.89
Output order	-0.02	1.00	0.05	0.99

predicted recall: a 1-SD increase in associative strength increased the odds of cue selection by 1.27 (95% CI: [1.17, 1.38]), a 1-SD increase in cue commonality by 1.47 times (95% CI: [1.33, 1.63]), and a 1-SD decrease in the number of associates (i.e., a more distinctive cue) increased the odds of selection by 1.44 times (95% CI: [1.33, 1.57]). Further, the magnitude of these influences on cue selection were roughly similar to their influences on actual recall.

In addition, however, participants' selections were influenced by one feature that had no bearing on actual recall: output order. A 1-standard increase in output order – that is, a cue output about 1.85 serial positions later – decreased the odds that a cue would be selected by 1.89 times (95% CI: [1.73, 2.08]). In fact, this effect was of larger magnitude than any other predictor, even though output order was *not* significantly predictive of actual recall.

Discussion

Honoring participants' choices about which self-generated cues to present at test produced better memory than dishonoring their choices. Given the general mnemonic

benefits of generation (Slamecka & Graf, 1978), the mere act of generating mnemonic cues could be expected to benefit memory through deeper processing of the targets and/or of the relationships between cues and targets (Fisher & Craik, 1980). Although generation may indeed help recall, Experiment 1 shows that specific self-generated cues have mnemonic benefits over and above generation. Both the Honor and Dishonor conditions involved a cue-generation process, but getting the *particular* cues requested in the Honor condition resulted in superior memory. Thus, Experiment 1 implies that learners can select self-generated memory cues that are likely to be especially effective. Requiring participants to output multiple potential mnemonic cues per target allows us to examine the processes of cue generation and selection separately, which no prior research has done. The results suggest that cue selection contributes to the mnemonic benefits of self-generated cues.

Table 3 Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 1 as a function of cue characteristics

Fixed effect	β	SE	Wald z	p
Intercept (baseline log odds of recall)	1.064	0.133	7.99	< .001
Cue-to-target associative strength	0.268	0.091	2.94	< .001
Number of associates	-0.416	0.067	-6.18	< .001
Cue commonality	0.374	0.084	4.47	< .001
Output order	-0.043	0.065	-0.67	.50

Note. SE = standard error

Table 4 Means and standard deviations in Experiment 1a of cue characteristics for cues that were and that were not selected measured in original units (top half) or standardized scores (bottom half)

Cue characteristic	Cues selected		Cues not selected	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cue-to-target associative strength	0.05	0.11	0.02	0.07
Number of associates	8.69	7.69	10.02	7.69
Cue commonality	18%	22%	11%	17%
Output order	2.09	1.43	3.12	1.91
Cue characteristic (<i>z</i> -scored)				
Cue-to-target associative strength	0.24	1.30	-0.09	0.84
Number of associates	-0.12	1.00	0.05	1.00
Cue commonality	0.28	1.17	-0.11	0.90
Output order	-0.41	0.78	0.16	1.03

Table 5 Fixed effect estimates for mixed effects logit model of cue choice in [Experiment 1a](#) as a function of cue characteristics

Fixed effect	β	SE	Wald z	p
Intercept (baseline log odds of choice)	-1.142	0.061	-18.82	< .001
Cue-to-target associative strength	0.239	0.041	5.81	< .001
Number of associates	-0.366	0.043	-8.52	< .001
Cue commonality	0.389	0.051	7.60	< .001
Output order	-0.638	0.047	-13.54	< .001

Note. SE = standard error

The benefit of self-selected cues over non-selected cues is obtained across all output positions.

Further, learners selected fewer than half of the first cues as their chosen mnemonic cue; the percentage of times that learners selected the first cue ranged across participants from 20% to 72%, with a grand mean of 44%. These results indicate that learners do not exclusively rely on the first mnemonic cue that comes to mind to serve as their final cue but rather deliberate over multiple cues before selecting one.

Why did honoring participants' cue selections improve memory? One reason is likely that these selections were attuned to several properties that made cues effective at cuing the target. Cue-target pairs were more apt to be recalled when the cue words were distinctive, commonly generated, and strongly associated with the target, and participants' cue selections were appropriately sensitive to all three of these properties. Participants also strongly favored cues that they output earlier in the generation process, even though output order had no actual relation to the usefulness of a cue; we revisit this pattern in the [General discussion](#).

So far, we have demonstrated that learners can choose cues that support their memory more effectively than the average cue. But for cue selection to contribute to the benefits of self-generated memory cues over other-generated cues, the cues that learners select for *themselves* must be better than the cues that others would select for them.

Experiment 1b

In [Experiment 1b](#), we tested an additional prediction of our cue-selection hypothesis: People should be better at selecting memory cues for themselves than for other learners. Thus, we examined how – and how effectively – people would choose cues for another, different learner. New participants (*observers*) saw the list of potential cues that a yoked participant generated in Experiment 1 and tried to select the most effective cues for the cue generator – first by selecting among all of the generated cues, and then by choosing between just the learners' preferred cue and its control from the dishonor

condition. In [Experiment 1a](#), honored cues more effectively supported recall than dishonored cues. Thus, if observers select the dishonored cues, that would be evidence that people are better selecting a cue for themselves than for others.

Further, to the extent that observers and generators differ in the efficacy of their cue selections, we can ask whether these differences in preferences can be accounted for the characteristics of these cues they select. In [Experiment 1a](#), we found that cue generators favored distinctive, common, and strongly associated cues; are observers sensitive to these properties?

Method

Participants As in [Experiment 1a](#), 34 participants completed the experiment for partial course credit in introductory educational psychology classes.

Procedure Each participant in this experiment was yoked to a single participant in [Experiment 1a](#). These participants were instructed about the details of the prior experiment. We told them that prior participants had generated several potential cues to help support their memories for target words on a later cued-recall test. The new participants were instructed to choose the most helpful cue for the prior participant. This choice was made in two steps that were designed for comparison with the choices made by the original generators. First, participants saw the target word displayed at the top of the screen. All of the potential cue words that a specific prior participant generated were displayed at the bottom of the screen in the order in which generators output them. Participants selected the cue word that they believed would be most beneficial for the prior participant. This choice emulated the selection by the generators among all of their cues.

If the prior participant generated more than two possible cues, the participant completed an additional step for the target. Namely, after they selected one out of all of the possible cues, the possible cues were narrowed down to two possible cues: the one that the prior participant chose (the one displayed in the honor condition) and a random non-selected cue (the one that *would have been* used in the dishonor condition). Participants then selected which cue out of these two would be more effective. This step was included to provide a more direct analogue to the experimental contrast between the honor and dishonor conditions in Experiment 1. Further, reducing the number of possible cue options should simplify the choice and, therefore, provide a stronger test of the cue-selection hypothesis. For example, if observers fail to choose effectively even when presented with only two options, their poor choices cannot be attributed to choice overload (e.g., Lee & Lee, 2004).

Results

Similarity of cue choice We first calculated how frequently the yoked participant in [Experiment 1b](#) picked the same cue as the generators in [Experiment 1a](#). Observers chose the same cue as the generators for 46% ($SD = 10\%$) of the targets. We compared this observed level of agreement to the proportion of agreement that would be obtained by choosing randomly out of the list of potential cues ($M = 32\%$, $SD = 9\%$). Observers chose the same cue as generators more frequently than random selections, $t(33) = 8.78$, $p < 0.001$, $d = 1.52$. In fact, only two of 34 participants chose the generator-selected cues less frequently than random selections. The distribution of observers' choices by output position is displayed in [Fig. 2](#).

We similarly compared how often the observer selected the honor choice between the two possibilities of the honor cue and the dishonor cue. Observers selected the honor choice for 61% ($SD = 9\%$) of the trials, which is significantly more frequently than chance (50%; $t(33) = 7.06$, $p < 0.001$, $d = 1.21$), but also significantly less than perfect agreement (100%; $t(33) = 25.38$, $p < .001$, $d = 4.35$).

Relation of cue characteristics to selection The above analyses indicated that observers showed some, though not perfect, agreement with generators' cue choices. Did this level of agreement obtain because of the degree to which observers did – or did not – base their selections on the same [cue characteristics](#) that generators did? [Table 6](#) displays the average characteristics of cues selected by the observer versus those not selected.

[Table 7](#) displays the results of a mixed-effects model of the (log) odds that an observer would select a cue as a function of that cue's characteristics. Observers were sensitive to many of the same characteristics as generators: a 1-SD increase in associative strength increased the odds of cue selection by 1.12

Table 6 Means and standard deviations in [Experiment 1b](#) of cue characteristics for cues that were and that were not selected measured in original units (top half) or standardized scores (bottom half)

Cue characteristic	Cues selected		Cues not selected	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cue-to-target associative strength	0.05	0.11	0.02	0.07
Number of associates	10.55	7.53	9.31	7.75
Cue commonality	22%	23%	9%	15%
Output order	1.92	1.26	3.18	1.91
Cue characteristic (<i>z</i> -scored)				
Cue-to-target associative strength	0.24	1.26	-0.09	0.86
Number of associates	0.12	0.98	-0.04	1.01
Cue commonality	0.49	1.24	-0.19	0.82
Output order	-0.49	0.69	0.19	1.04

Table 7 Fixed effect estimates for mixed effects logit model of cue choice in [Experiment 2](#) as a function of cue characteristics

Fixed effect	β	SE	Wald z	p
Intercept (baseline log odds of choice)	-1.223	0.065	-18.99	< .001
Cue-to-target associative strength	0.113	0.036	3.14	< .001
Number of associates	0.007	0.038	0.16	.88
Cue commonality	0.591	0.047	12.72	< .001
Output order	-0.807	0.051	-16.09	< .001

Note. SE = standard error

(95% CI: [1.04, 1.20]), a 1-SD increase in cue commonality by 1.81 times (95% CI: [1.65, 1.98]), and a 1-SD increase in output order decreased the odds of selection by 2.24 times (95% CI: [2.03, 2.47]). Notably, however, and unlike cue generators, there was *no* significant effect of cue distinctiveness; in fact, cue generators numerically favored cues that were *less* distinctive (i.e., had more associates).

Next, we directly compared whether generators in [Experiment 1a](#) based their selections on different cue characteristics than observers in [Experiment 1b](#). Using only the trials where the generator and observer chose different cues, we determined which characteristics predicted a cue being selected by the generator compared to the yoked observer. [Table 8](#) displays the results of this model, which explicitly tests the differences between [Table 5](#) and [Table 7](#). Generators and observers differed in their use of each of the cue characteristics. More specifically, generators' selections were more sensitive to cue-to-target associative strength, were more sensitive to the number of associates, were less sensitive to cue commonality, and were more sensitive to output order than observers' selections were.

Discussion

In [Experiment 1b](#), we tested whether participants could select cues for another learner as effectively as the learners who had generated those cues for themselves. Observers showed an intermediate degree of agreement with the generators' choices; they selected the generator's choices more frequently than chance, but they showed significantly less than perfect alignment with the generator. This held true both for the choice among all the generators' generated cues as well as a choice specifically between the generator's chosen cue and its dishonor control. This latter comparison is important because it implies that observers not only chose *different* cues than the original learners, but – given that the honor cues were on average better than the dishonor cues – they also chose *less-effective* cues (a point we test further in [Experiment 2](#)).

Table 8 Fixed effect estimates for mixed effects logit model of self-selected cues relative to observer-selected cues across Experiments 1a and 1b as a function of cue characteristics

Fixed effect	β	SE	Wald z	p
Intercept (baseline odds of generator selection)	0.034	0.054	0.63	.532
Cue-to-target associative strength	0.162	0.057	2.83	.005
Number of associates	-0.416	0.058	-7.14	< .001
Cue commonality	-0.254	0.063	-4.03	< .001
Output order	0.165	0.052	3.15	.002

Note. SE = standard error

Why could people better select cue for themselves than for others? Analysis of the cue characteristics that predicted observers' cue selections suggests that observers utilized cue characteristics differently than generators did when selecting cues for themselves. Generators relied upon cue-to-target associative strength, cue distinctiveness, and output order more than observers, while observers relied more upon cue commonality. Greater reliance on cue commonality may indicate that observers do not appreciate that idiosyncratic cues (which could tie into strong personal experiences). Critically, while generators valued distinctive cues, observers picked cues that pointed to more potential targets. This failure to capitalize on distinctiveness likely damages observers' cue selections given that our analysis of cue diagnosticity indicated that cue distinctiveness was the feature most strongly predictive of recall (consistent with past work; Nairne, 2002).

Experiment 2

Comparing the data across Experiments 1a and 1b, we can infer that cue selections by others are less effective than cue selections by oneself. Yet, we could not directly compare the efficacy of cues selected by oneself and others because we could not give observer-selected cues to the cue generator, who had already completed the experiment days or weeks earlier. In this final experiment, learners simultaneously participated in dyads. The cues present at the final cued-recall test included a generator's selections, the paired observer's selections, and dishonored selections. The addition of observer-selected cues in the recall test allows us to directly compare how effectively cues selected by oneself support retrieval compared to cues selected by others.

Method

Participants As in both prior experiments, 34 participants completed the experiment for partial course credit in introductory educational psychology classes.

Procedure Participants signed up to complete the experiment through a participant-management system online. If two participants showed up for the same session, they completed this particular experiment. Otherwise, they completed an unrelated experiment on their own. Participants were given the same instructions as in Experiment 1 about creating cues to support their memory retrieval later. Three changes were made from Experiment 1. First, participants had to supply at least three cues for each target. Second, two separate lists of 45 different targets were used. Each list had 25 targets from Experiment 1 and 20 additional new targets. The two lists did not share any target items. Two lists were needed because a distinct list was given to each participant in a pair. The third change was that participants chose cues for their partner after they finished generating cues for their list of 45 targets. This cue selection phase mirrored that from Experiment 1b closely. Participants were told that they would see the list of potential cues that another participant had generated and were asked to select the most effective cue for the other participant. Participants were not told that their partner participant was also present in the lab. Participants saw the target with the list of the cues that their partner generated in the order that their partner generated them and typed in their selection. Whenever one participant in a pair finished a phase before their partner, they played a computer puzzle game in the NetLogo programming library called Planarity, in which participants had to untangle a digital knot by moving connected nodes.

After both partners finished selecting cues for each other, their memories for the targets were tested. Targets were randomly (but equally) assigned within-subjects to three conditions: self-selected cue, observer-selected cue, or dishonored cue. In the self-selected cue condition, participants received the cue that they selected. In the observer-selected cue condition, participants received the cue that their partner selected for them. Finally, in the dishonor condition, participants received a random cue from the list that they generated, as long as it was not the cue that the participants selected for themselves.

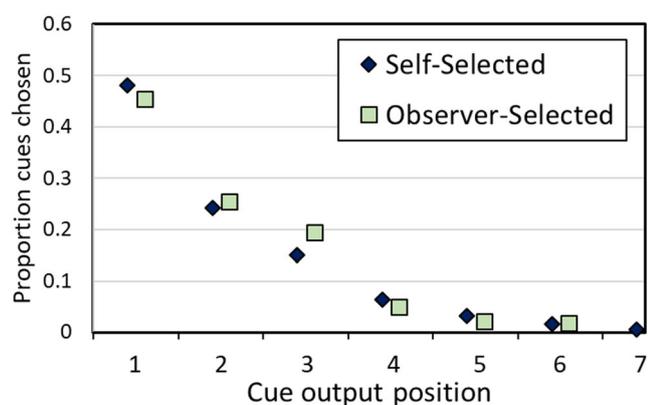


Fig. 4 Proportion of cues chosen by output position by the cue generator and by the observer. Data points are jittered horizontally to display each clearly

Results

Similarity of cue choice We first calculated the overlap between generators' and observers' cue selections (see Fig. 4). Regardless of output order, observers selected the same cues as the generators ($M = .49$, $SD = .11$) more than would be expected if they were randomly choosing one of the supplied cues ($M = .28$, $SD = .05$), $t(33) = 11.37$, $p < .001$, $d = 1.98$, indicating significant overlap between generators' and observers' selections. Only one of 34 participants chose the generator-selected cues less frequently than random selections.

Effectiveness of cue choice Next, we examined how recall differed between self-selected cues, observer-selected cues, and random cues. The average recall by output position and condition is shown in Fig. 5. To compare these three types of trials in our mixed logit model, we used two orthogonal contrasts. The first compared the self- and observer-selected cues to random cues to assess the benefits of human-selected cues relative to random ones. The second contrast specifically compared self-selected to observer-selected cues to determine

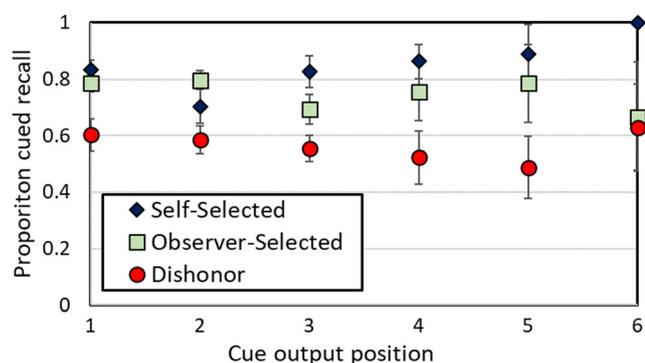


Fig. 5 Proportion of correct cued recall as a function of cue condition and cue output position in Experiment 2. Error bars show standard errors of the mean; error bars get progressively wider with output position due to the reduced number of cues supplied at those positions

Table 9 Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 2 as a function of condition

Fixed effect	$\hat{\beta}$	SE	Wald z	p
Intercept (baseline log odds of recall)	1.151	0.161	7.17	< .001
Selected cues (vs. random)	1.228	0.173	7.08	< .001
Self-selected (vs. observer-selected)	0.440	0.172	2.56	.010
Output order	-0.040	0.053	-0.75	.455
Selected cues \times Output order	0.014	0.107	0.13	.893
Self-selected \times Output order	0.097	0.135	0.72	.474

Note. SE = standard error

whether participants could more effectively select a cue for themselves than for another person.

Table 9 displays the results of this model.³ Both contrasts were significant: When given a cue that someone had selected, the odds of correct recall ($M = 79\%$) were 3.41 (95% CI: [2.43, 4.79]) times greater than when given a random cue ($M = 56\%$), $z = 7.08$, $p < 0.001$. There was also a smaller, but significant, benefit of the self-selected cues ($M = 82\%$), with the odds of correct recall being 1.55 (95% CI: [1.11, 2.17]) over the observer-selected cues ($M = 76\%$), $z = 2.56$, $p = .01$.

Relation of cue characteristics to recall Thus far, we have seen that (a) cues selected by either the generator or the observer were substantially better than random cues, and (b) self-selected cues were more effective than observer-selected cues. Can the characteristics of the selected cues explain both of these effects?

We first examined cue diagnosticity by examining which cue characteristics predicted recall. Table 10 displays the mean cue characteristics for cue-target pairs that were recalled versus those that were not, and Table 11 shows the results of the mixed logit regression. The results generally mirrored those in Experiment 1a: All three cue characteristics positively predicted recall even controlling for output order, with the strongest effect being that of cue distinctiveness. However, one difference was that, in Experiment 2, output order also predicted performance above and beyond the three beneficial cue characteristics, with the odds of recall declining by 1.19 times (95% CI: [1.05, 1.36]) for each subsequent serial position at which the cue was output. This suggests that later-generated cues may unfavorably differ in other properties, beyond distinctiveness, associative strength, and commonality, that affect recall.

³ On 49% of trials, the self-selected and partner-selected were the same. These trials necessarily diminish any difference between self-selected and partner-selected cues since, when the cues are the same, it is not possible for accuracy to be systematically greater in one condition than the other. Nevertheless, the analysis reported in the main text observed a significant benefit of self-selected cues even with the inclusion of these trials, and when we performed an analysis on just the trials where the self- and partner-selected cues differ, the difference was even larger (82% vs. 69%, $z = 3.39$, $p = .001$.)

Table 10 Means and standard deviations in [Experiment 2](#) of cue characteristics for cue-target pairs that were and that were not remembered, measured in original units (top half) or standardized scores (bottom half)

Cue characteristic	Items remembered		Items not remembered	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cue-to-target associative strength	0.04	0.11	0.01	0.05
Number of associates	9.49	7.45	11.71	7.11
Cue commonality	9%	11%	7%	10%
Output order	2.11	1.30	2.37	1.35
Cue characteristic (<i>z</i> -scored)				
Cue-to-target associative strength	0.08	1.12	-0.20	0.55
Number of associates	-0.09	1.00	0.21	0.96
Cue commonality	0.06	1.03	-0.16	0.89
Output order	-0.06	0.99	0.14	1.02

Table 11 Fixed effect estimates for mixed effects logit model of cued recall accuracy in [Experiment 2](#) as a function of cue characteristics

Fixed effect	$\hat{\beta}$	SE	Wald <i>z</i>	<i>p</i>
Intercept (baseline log odds of recall)	1.223	0.156	7.85	< .001
Cue-to-target associative strength	0.405	0.116	3.52	< .001
Number of associates	-0.459	0.074	-6.22	< .001
Cue commonality	0.334	0.084	4.00	< .001
Output order	-0.177	0.067	-2.66	.01

Note. SE = standard error

Relation of cue characteristics to selection Can the differential effectiveness of self-selected, observer-selected, and random cues be explained by the characteristics of the cues? For [Experiment 2](#), we examined cue utilization both for self-selected cues (left panel of [Table 12](#)) and observer-selected cues (right panel of [Table 12](#)).

We ran two logit mixed models to predict cue choice as a function of cue characteristics: One predicted learners' choices for themselves and one predicted their partners' choices. [Table 13](#) displays the results of these models. Both learners' choice and partners' choices were sensitive to all three of the beneficial cue characteristics reviewed above: cue-to-target associative strength, cue distinctiveness⁴ (i.e., a

⁴ This effect of cue distinctiveness diverges from the bivariate relationship presented in [Table 12](#), where selected cues if anything had more associates. This apparent discrepancy is explained by the fact that the model results in [Table 13](#) present the effect of cue distinctiveness *while controlling for the other, potentially confounding cue characteristics*. In particular, the number of associates had a modest positive relationship with cue commonality ($r = .29$), such that more commonly provided cues also tend to have more associates (i.e., are less distinctive). Given that participants tended to select more common cues, on average, chosen cues had more associates than unchosen cues. However, the multiple regression presented in [Table 13](#) demonstrates that, when controlling for cue commonality, participants tended to select cues with *fewer* associates (i.e., the relationship reverses once the confounder of cue commonality is properly controlled for; Simpson, 1951).

low number of associates), and cue commonality. In addition, both choices were predicted by output order; this relationship is rational given that output order did predict recall in [Experiment 2](#). Overall, then, these results suggest that learners and partners were sensitive to cue characteristics predictive of recall, which may explain at least in part why both self-selected and partner-selected cues were better than random cues.

Next, we examined the differences between self-chosen and observer-chosen cues. Similar to our analysis of [Experiment 1b](#), we took the 785 trials where the learner and partner chose different cues and used another mixed logit model to determine which features were predictive of a cue being chosen by the learner as opposed to the observer. [Table 14](#) displays the results of this model. The only cue characteristic that reliably differed between self- and partner-chosen cues was cue commonality: An observer's choice of cue was more influenced by cue commonality than was the learner's own choice, replicating the difference observed in [Experiment 1](#). (Learners were also numerically more sensitive to cue distinctiveness, as in [Experiment 1](#), but this effect did not reach conventional levels of significance in [Experiment 2](#).) Overall, this analysis suggests that the benefits of self-chosen cues did not arise from greater sensitivity to normatively desirable cue characteristics because observer-chosen cues were not significantly less sensitive to these characteristics (and, indeed, were *more* sensitive to cue commonality). Instead, the comparatively small difference between self-chosen and partner-chosen cues may reflect learners capitalizing on more idiosyncratic features that partners would not know about.

Discussion

[Experiment 2](#) provided a particularly strong test of differences in the mnemonic cues selected for oneself versus others because observers experienced the process of cue generation and

Table 12 Means and standard deviations in [Experiment 2](#) of cue characteristics for cues that were and that were not selected measured in original units (top half) or standardized scores (bottom half)

Cue characteristic	Self-selected				Observer-selected			
	Cues selected		Cues not selected		Cues selected		Cues not selected	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Associative strength	0.07	0.14	0.03	0.09	0.05	0.12	0.02	0.07
Number of associates	10.29	7.12	10.10	7.48	10.17	7.41	10.09	7.44
Cue commonality	14%	12%	8%	10%	12%	12%	6%	8%
Output order	1.73	1.13	2.55	1.35	2.04	1.22	2.74	1.38
Cue characteristic (<i>z</i> -scored)								
Associative strength	0.41	1.40	-0.08	0.89	0.21	1.24	-0.18	0.69
Number of associates	0.02	0.96	0.00	1.01	0.01	1.00	0.00	1.00
Cue commonality	0.50	1.15	-0.09	0.94	0.32	1.12	-0.27	0.79
Output order	-0.34	0.86	0.06	1.01	-0.12	0.90	0.10	1.07

selection themselves before selecting cues for others. This design isolates the effect of selecting cues for oneself while holding constant experience with generating and selecting cues.

This experiment replicated the finding from Experiment 1 that self-selected mnemonic cues supported cued recall better than random cues (i.e., dishonored selections), and it provided direct evidence that cues that are both self-generated *and* self-selected are better than cues generated by oneself but selected by an observer. Generators may value idiosyncratic cues because they understand their unique knowledge and processing, while observers have no knowledge of the generator's knowledge. Additionally, while observers in [Experiment 1b](#) did not value the distinctiveness of cues, observers in [Experiment 2](#) showed some appreciation of distinctiveness. This may be because observers in [Experiment 2](#) had previously generated and selected cues for themselves. Generating and selecting cues for oneself may highlight the need for

distinctive cues and cause learners to see the importance of this cue characteristic for themselves and others.

General discussion

Across two experiments, we tested whether the ability to select effective cues contributes to the benefits of creating one's own mnemonic cues. Requiring participants to output multiple potential mnemonic cues per target distinguishes the impact of cue selection from cue generation on later cued recall. Our research is the first to analyze the generation and selection of mnemonic cues separately and shows that the process of selecting cues contributes to the benefits of self-generated cues. While not excluding potential direct benefits of generation to the strength of memory traces, our results suggest that the benefits of self-generated cues are at least in part metamnemonic. When considering mnemonic cues, learners

Table 13 Fixed effect estimates for mixed effects logit model of cue choice in [Experiment 2](#) as a function of cue characteristics for self-selected cues (left panel) and observer-selected cues (right panel)

Fixed effect	Self-selected				Observer-selected			
	$\hat{\beta}$	<i>SE</i>	Wald <i>z</i>	<i>p</i>	$\hat{\beta}$	<i>SE</i>	Wald <i>z</i>	<i>p</i>
Intercept (baseline odds of choice)	-1.378	0.046	-29.74	< .001	-1.382	0.045	-30.78	< .001
Cue-to-target associative strength	0.243	0.038	6.46	< .001	0.233	0.036	6.38	< .001
Number of associates	-0.297	0.042	-7.05	< .001	-0.211	0.041	-5.13	< .001
Cue commonality	0.413	0.045	9.11	< .001	0.618	0.045	13.75	< .001
Output order	-0.783	0.050	-15.75	< .001	-0.717	0.049	-14.63	< .001

Note. SE = standard error

Table 14 Fixed effect estimates for mixed effects logit model of self-selected cues relative to observer-selected cues in [Experiment 2](#) as a function of cue characteristics

Fixed effect	β	SE	Wald z	p
Intercept (baseline odds of learner choice)	0.037	0.055	0.67	.502
Cue-to-target associative strength	0.009	0.043	0.21	.833
Number of associates	-0.089	0.053	-1.67	.095
Cue commonality	-0.259	0.051	-5.10	< .001
Output order	-0.050	0.063	-0.79	.430

Note. SE = standard error

typically output more than two possible choices and selected the first cue that came to mind only half of the time. This suggests that cue generation often involves a process of selection among multiple potential cues. Further, subsequent cued recall was consistently greater when learners' cue selections were honored than dishonored, suggesting that learners could indeed identify and select particularly effective memory cues, even among a list of cues that they generated. Finally, generators selected more effective memory cues for themselves than observers could, indicating that this ability was at least partially specific to selecting cues for one's *own* memory.

These differences in cue effectiveness can be explained at least partially by systematic differences in the kinds of cues that people chose. The cue selection criteria used by generators and observers somewhat overlap but differ in important ways. Specifically, generators and observers both appreciated normative cue-to-target associative strength, but generators valued distinctiveness more than observers and observers valued cue commonality more than generators. Our data also suggest that the cue characteristics carry import even when controlling for the impact of choice on recall.

An alternative explanation of our results is that the act of selecting cues itself facilitated memory for those items; recent research suggests that selected items are more memorable than unselected items (Coverdale & Nairne, 2019). However, in the [Appendix](#), we report a supplementary analysis examining whether recall accuracy could be predicted both from our existing cue characteristics and from a variable capturing whether the cue was selected or not. The addition of the selection variable did not eliminate the impact of cue characteristics on recall. In other words, cue characteristics predicted recall, even when accounting for the beneficial effects of choice on recall.

Perspective-taking in memory

One reason that people were more effective in selecting cues for their own memory is that cue generators valued the distinctiveness of a cue more than observers. Observers' failure to value cues' distinctiveness as much as generators do echoes

prior research (Tullis & Fraundorf, 2017) and may reveal a limitation of our ability to take the perspective of others (Keysar et al., 2003). The current experiments suggest that, even when distinctive cues that are generated by others are available for selections, observers fail to value distinctiveness of cues as much as generators. Although critical to retrieval (Nairne, 2002), distinctiveness may be a particularly difficult cue to utilize when selecting cues because its benefits may be only apparent during retrieval (i.e., when the target is absent). During encoding and selection, learners see both the cue and target, which can allow learners to easily notice and judge the cue-to-target associative strength. However, the presence of the target during cue generation may obscure considerations for how much the cue narrows down the field of potential targets during retrieval. Generators, then, may value this characteristic more than observers because generators see the target without cues, while observers never study the target without the presence of the cue. The underappreciation of distinctiveness may suggest that observers have a poor model of the needs of memory. But, experience with the task may alleviate some of this perspective-taking burden. When observers experienced the task themselves in [Experiment 2](#), their cue selections became somewhat sensitive to distinctiveness.

By comparison, observers value cue commonality *more* than generators, and the difference in the value of cue commonality likely reveals an additional challenge to taking perspective. Observers likely chose more common cues than generators because they did not recognize how unique cues would fit with someone else's experiences. Even though observers were shown all the possible cues for the generator, including idiosyncratic cues, they favored more generic common cues – and favored generic cues to a greater extent than generators did. Differences in the value placed upon these cue characteristics by generators and observers reveals a limitation in our ability to effectively take the perspective of others.

Output order

Beyond the beneficial effects of associative strength, cue distinctiveness, and cue commonality, participants' cue

selections were associated with at least one other variable: output order. In both experiments, both generators and observers tended to favor earlier-output cues, even when controlling for the other cue characteristics. In at least some cases, this preference may be rational: In [Experiment 2](#), earlier output order positively predicted recall over and above the cue characteristics. Although it is possible that reflects a causal effect of output order itself on recall, it is perhaps more plausible that this reflects a correlation between output order and some other, unmeasured cue characteristic – for example, imageability or frequency – that benefits recall and correlates with output order. In other cases, namely [Experiment 1](#), participants preferred earlier cues even when output order was fully uncorrelated with recall (see also [Tullis & Fraundorf, 2017](#)). This tendency may thus reflect the broader trend for order to influence or bias decision-making, including in the domain of self-regulated learning (e.g., [Ariel et al., 2011](#)).

Beliefs versus experience

Although we found that both generators and observers' choices were sensitive to a broad palette of cue characteristics – some more than others – it is presently unclear whether these choices represent a form explicit or implicit knowledge. Prior work (e.g., [Fraundorf & Benjamin, 2014](#); [Kelley & Jacoby, 1996](#); [Koriat et al., 2004](#)) has established that metacognitive judgments are sometimes made based on learners' explicit, verbalizable beliefs but sometimes based on their in-the-moment experience with individual memoranda, which do not necessarily align. For example, learners may have favored distinctive cues because of a naïve theory they held about what constitutes a good memory cue. But, they could have also selected these cues for other reasons: they resembled cues they have used in the past, for instance, or simply “felt right.” One way for future work to probe learners' conscious beliefs about effective cues is to adopt a method in which participants read a description of the experiment and judge performance without personally experiencing any individual stimulus (e.g., [Koriat et al., 2004](#); [Kornell et al., 2011](#)).

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). 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., [Fraundorf & Benjamin, 2016](#); [Finley & Benjamin, 2021](#)). However, effectively anticipating and controlling

available retrieval routes (i.e., retrieval environments) during encoding may be an especially challenging form of self-regulated learning because learners must take the perspective of their selves in the future (e.g., [Kornell & Bjork, 2009](#)). Recent experiments have yielded conflicting results about how effectively learners exercise control over their later retrieval contexts, with some research showing that learners effectively choose test questions that match how they encoded the information ([Finley et al., 2012](#)), but other work showing a lack of such appreciation ([Finley & Benjamin, 2019](#)).

In the context of cue generation, learners must predict whether the cue they generate will lead them to the appropriate target during later retrieval (i.e., whether they will be able to decode their mediator; [Pyc & Rawson, 2010](#)). 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](#)). Mental states naturally shift over time due to new experiences and development (e.g., [Estes, 1955](#)). If learners cannot accurately anticipate these changes when generating mnemonic cues, the effectiveness of cues may suffer. Nevertheless, despite natural shifts in our mental contexts (e.g., [Kornell & Bjork, 2009](#)), the present results suggest that generators can select cues that are more effective for themselves than others can. This provides even more evidence of the importance of allowing learners to exercise some control over their own learning ([Tullis & Benjamin, 2011](#)).

Self-generated cues allow learners to remember information efficiently and effectively. Nevertheless, while self-generated cues are more effective than other-generated cues, they do not yield perfect recall. Self-generated memory cues can fail in two ways: A learner can fail to retrieve the cue (i.e., retrieval deficiency) or the learner can forget how to interpret the cue (i.e., decoding deficiency; [Dunlosky et al., 2005](#), [Tullis & Qiu, in press](#)). In our experiments, learners were provided with their mnemonic cue during recall. Therefore, 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](#)). Future studies can examine how generating cues impacts both retrieval deficiency and cue decodability.

Applications

People generate cues to help them remember information when they take notes in meetings, write to-do lists, name computer files, learn prescription schedules, and encode new acquaintances' names ([Tullis & Finley, 2018](#)). Additionally, students consistently report generating mnemonics to support

their learning of new information across a variety of content domains (e.g., Tullis & Maddox, 2020). Successful memory across these diverse tasks may depend upon the creation and selection of effective mnemonic cues. Our results reinforce prior results showing that learners can appropriately generate cues to help their memories (e.g., Tullis & Fraundorf, 2017) and broaden the growing metacognitive literature that suggests that learners can effectively control their own learning, including allocating study time (Tullis & Benjamin, 2011), scheduling study (Benjamin & Bird, 2006), and choosing study strategies (Tullis et al., 2018).

Critically, our current results suggest that people can choose effective cues over ineffective cues, even among those that they generate themselves, and that the benefits of self-generated cues indeed come largely from appropriate *selections* among multiple cue candidates (rather than from the process of *generation* itself). Allowing learners to select a specific memory cue for themselves among a list of provided options may be less cognitively demanding than generating one's own (and require less time) and may not sacrifice mnemonic efficacy. Consequently, teachers or textbooks could provide multiple mnemonics and students could individually choose cues that are effective for their own personal memories. Further, by providing for the selection of personal mnemonic cues among potential cue candidates, we may be able to support memory among low-performing students, who spontaneously generate fewer mnemonic cues than others (Scruggs et al., 1986), students with learning disabilities (Mastropieri & Scruggs, 2000), and elderly adults, who may struggle to generate their own cues (Verhaeghen et al., 1992). Understanding the processes, strengths, and weaknesses of cue generation may ultimately allow us to develop targeted interventions to help people utilize better cues so that they can more effectively remember and apply information.

Data Availability None of the experiments was preregistered. The data and materials for all experiments are available at https://osf.io/muhtc/?view_only=36d20866557444fb914c3e26784faf0

Appendix

We conducted additional analyses predicting recall from cue characteristics and whether the generator selected it. As shown in the tables below, the cue characteristics retain their import, even after accounting for whether the item was selected or not. Further, even when accounting for the four cue characteristics listed, selected cues yielded better recall than unselected cues. The mnemonic benefits for selected cues over unselected cues may be driven by choice itself (e.g. Coverdale & Nairne, 2019) or by other, unmeasured cue characteristics that differ between selected and unselected cues.

Table 15 Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 1a as a function of cue characteristics and generator choice

Fixed effect	β	SE	Wald z	p
Intercept (baseline odds of recall)	0.492	0.145	3.39	< .001
Cue-to-target associative strength	0.215	0.087	2.47	.01
Number of associates	-0.351	0.068	-5.14	< .001
Cue commonality	0.289	0.084	3.45	< .001
Output order	-0.040	0.068	-0.59	.56
Chosen by generator	1.225	0.132	9.31	< .001

Note. SE = standard error

Table 16 Fixed effect estimates for mixed effects logit model of cued recall accuracy in Experiment 2 as a function of cue characteristics and generator choice

Fixed effect	$\hat{\beta}$	SE	Wald z	p
Intercept (baseline odds of recall)	1.125	0.159	7.07	< .001
Cue-to-target associative strength	0.361	0.113	3.21	.001
Number of associates	-0.446	0.074	-6.07	< .001
Cue commonality	0.284	0.084	3.38	< .001
Output order	-0.155	0.067	-2.32	.02
Chosen by generator	0.636	0.215	2.96	.003

Note. SE = standard error

References

- Ariel, R., Al-Harthi, I. S., Was, C. A., & Dunlosky, J. (2011). Habitual reading biases in the allocation of study time. *Psychonomic Bulletin & Review*, 18, 1015-1021.
- 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*, 104(2), 126-13.
- Bellezza, F. S., & Poplawsky, A. J. (1974). The function of one-word mediators in the recall of word pairs. *Memory & Cognition*, 2, 447-452.
- Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language*, 55(1), 126-137. <https://doi.org/10.1016/j.jml.2006.02.003>
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35, 201-210.
- 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.
- Coverdale, M. E., & Nairne, J. S. (2019). The mnemonic effect of choice. *Psychonomic Bulletin & Review*, 26, 1310-1316.
- Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, 41, 389-400.

- Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 717-726.
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, 62, 369-377.
- Finley, J. R., & Benjamin, A. S. (2021). Metacognitive control of transfer-appropriate processing. Manuscript in preparation.
- Finley, J. R., & Benjamin, A. S. (2012). Adaptive changes in encoding strategy with experience: Evidence from the test expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 632-652.
- 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*. : Springer Science & Business Media.
- Fisher, R. P., & Craik, F. I. M. (1980). The effects of elaboration on recognition memory. *Memory & Cognition*, 8, 400-404.
- Fraundorf, S. H., & Benjamin, A. S. (2014). Knowing the crowd within: Metacognitive limits on combining multiple judgments. *Journal of Memory and Language*, 71, 17-38.
- Fraundorf, S. H., & Benjamin, A. S. (2016). Conflict and metacognitive control: The mismatch-monitoring hypothesis of how others' knowledge states affect recall. *Memory*, 24, 1108-1122.
- Harris, J. E. (1980). Memory aids people use: Two interview studies. *Memory & Cognition*, 8, 31-38.
- Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 484-494. <https://doi.org/10.1037/0278-7393.14.3.484>
- 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, 217-225.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 17, 649-667.
- Jamieson, D.G., & Schimpf, M.G. (1980). Self-generated images are more effective mnemonics. *Journal of Mental Imagery*, 4, 25-33.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469-486.
- Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, 35, 157-175.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25-41.
- Kibler, J. L., & Blick, K. A. (1972). Evaluation of experimenter-supplied and subject-originated first-letter mnemonics in a free-recall task. *Psychological Reports*, 30, 307-313.
- Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 923-954
- Koriat, A., Bjork, R. A., Sheffer, L., Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133(4), 643-656.
- Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, 138, 449-468.
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(3), 609-622. <https://doi.org/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.
- 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), 23-38
- Lee, B. K., & Lee, W. N. (2004). The effect of information overload on consumer choice quality in an on-line environment. *Psychology and Marketing*, 21, 159-83.
- Levin, M. E., & Levin, J. R. (1990). Scientific mnemonics: Methods for maximizing more than memory. *American Education Research Journal*, 27, 301-321.
- Lovelace, E. A. (1984). Metamemory: monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 75-766.
- 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, 303-313.
- 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, 502-509.
- Mastropieri, M. A., & Scruggs, T. E. (2000). The effectiveness of mnemonic instruction for students with learning and behavior problems: An update and research synthesis. *Journal of Behavioral Education*, 10, 163-173.
- McFarland, C. E., Jr., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning & Verbal Behavior*, 19, 210-225.
- Naime, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, 10, 389-395.
- 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/>
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330, 335.
- Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43, 103-121.
- 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.
- 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, 40-45.
- Ryskin, R., 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, 898-915.
- 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, 207-216.
- Scruggs, T. E., Mastropieri, M. A., Jorgensen, C., & Monson, J. (1986). Effective mnemonic strategies for gifted learners. *Journal for the Education of the Gifted*, 9, 105-121.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*, 13, 238-241.
- 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*, 371-394.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66-73.
- Tullis, J. G. & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, *64*, 109-118.
- Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review*, *19*, 743-749.
- Tullis, J. G., & Benjamin, A. S. (2015a). Cue generation: How learners flexibly support future retrieval. *Memory & Cognition*, *43*, 922-938.
- Tullis, J. G., & Benjamin, A. S. (2015b). Cuing others' memories. *Memory & Cognition*, *43*, 634-646.
- 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. <https://doi.org/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*, 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., & Maddox, G. (2020). Self-reported use of retrieval practice varies by age and domain. *Metacognition and Learning*, *15*, 129-154.
- Tullis, J. G. & Qiu, J. (in press). Generating mnemonics boosts recall of chemistry information. : *Applied*.
- Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, *71*, 673-679.
- Van Loon, M. H., de Bruin, A. B. H., van Gog, T., van Merriënboer, J. J. G., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica*, *151*, 143-154.
- Van Etten, S., Freebern, G., & Pressley, M. (1997). College students' beliefs about exam preparation. *Contemporary Educational Psychology*, *22*, 192-212.
- Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging*, *7*, 242-251.

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