Benefits of Accumulating Versus Diminishing Cues in Recall

5 authors, including:

Jason R Finley
Southern Illinois University Edwardsville
31 PUBLICATIONS   615 CITATIONS

SEE PROFILE

Matthew Jensen Hays
University of Southern California
18 PUBLICATIONS   655 CITATIONS

SEE PROFILE

Robert A. Bjork
University of California, Los Angeles
180 PUBLICATIONS   21,581 CITATIONS

SEE PROFILE

Nate Kornell
Williams College
65 PUBLICATIONS   6,034 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Retrieval Effort and Learning View project

Benefits of accumulating versus diminishing cues in recall

Jason R. Finley [a,*], Aaron S. Benjamin [a], Matthew J. Hays [b,1], Robert A. Bjork [b], Nate Kornell [b,c]

[a] Department of Psychology, University of Illinois at Urbana-Champaign, USA
[b] Department of Psychology, University of California, Los Angeles, USA
[c] Department of Psychology, Williams College, USA

ARTICLE INFO

ABSTRACT

Optimizing learning over multiple retrieval opportunities requires a joint consideration of both the probability and the mnemonic value of a successful retrieval. Previous research has addressed this trade-off by manipulating the schedule of practice trials, suggesting that a pattern of increasingly long lags—"expanding retrieval practice"—may keep retrievals successful while gradually increasing their mnemonic value (Landauer & Bjork, 1978). Here we explore the trade-off issue further using an analogous manipulation of cue informativeness. After being given an initial presentation of English–Iñupiaq word pairs, participants received practice trials across which letters of the target word were either accumulated (AC), diminished (DC), or always fully present. Diminishing cues yielded the highest performance on a final test of cued recall. Additional analyses suggest that AC practice promotes potent (effortful) retrieval at the cost of success, and DC practice promotes successful retrieval at the cost of potency. Experiment 2 revealed that the negative effects of AC practice can be partly ameliorated by providing feedback after each practice trial.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Effortful retrieval enhances long-term learning (Bjork, 1975; Gardiner, Craik, & Bleasdale, 1973; Glover, 1989; Pyc & Rawson, 2009). This principle underlies the concept of *desirable difficulties* (Bjork, 1994), whereby durable and flexible gains are thought to result from conditions that make learning effortful (Schmidt & Bjork, 1992). These conditions include spaced repetitions of learning events (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), interleaved practice of contextually interfering tasks (Shea & Morgan, 1979), a reduction in the frequency of feedback (Schmidt, 1991), and the use of tests as learning events (Roediger & Karpicke, 2006). But, in each of these cases, difficulties are only desirable to the extent that they are overcome.

That is, practice can be sub-optimal not only when conditions are too easy, but also when they are too hard. Thus, harnessing difficulties to enhance learning is a matter of determining and implementing the appropriate amount of retrieval difficulty for a given learner and a given set of materials.

Desirable difficulties in the spacing of practice

The manipulation of retrieval difficulty that has received the most empirical attention is the scheduling of repeated practice. It is well established that longer lags (i.e., more spacing) between practice trials can impair performance during training but enhance it at a delay (cf. Bahrick, 1979; Dempster, 1988; Greene, 2008; Schmidt & Bjork, 1992). When the practice trials are tests, rather than re-study opportunities, spacing must be calibrated with respect to forgetting in order to take advantages of desirable difficulty. If the lag between trials is too short, forgetting is minimal, retrieval is trivial, and the benefits of successful retrieval are small. However, if the lag between trials is

---

too long, forgetting is considerable, retrieval is unlikely, and the benefits of successful retrieval are unrealized (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). Thus, hitting the "sweet spot" on the trade-off between the amount of forgetting and the magnitude of the benefit requires considerable calibration, not just for the level of difficulty on each individual practice trial, but also for the overall difficulty schedule (e.g., increasing, decreasing, or constant).

A prediction of this view is that an optimally difficult spacing schedule must provide shorter lags early in practice when levels of learning are low, and longer lags later in practice when levels of learning are higher. This prediction was tested by Landauer and Bjork (1978), who demonstrated that a schedule of *expanding* spaced intervals produced superior final test performance when compared to uniformly spaced or contracting spaced intervals.

The logic behind an expanding schedule is compelling. The best time to retrieve an item—the sweet spot of maximal potency—is just before that item is forgotten (i.e., after the longest manageable interval). As an item becomes better learned because of each successive retrieval, the maximum interval at which the learner can successfully retrieve that item increases, moving the sweet spot further away in time. A schedule of expanding retrieval practice should provide the best chances of fitting this pattern, offering successive retrieval opportunities that are neither too easy (as they may be with later practice trials in contracting or uniform spacing) nor too hard (as they may be with the early trials of contracting or uniform spacing).

Many prior studies have examined the theoretical underpinnings and the applied potential of expanding retrieval practice (for reviews, see Balota, Duchek, & Logan, 2007; Storm, Bjork, & Storm, 2010), and Pavlik and Anderson (2008) developed a spacing optimization model that produced expanding schedules. It is important to note, however, that expanding schedules are not guaranteed to always optimally promote long-term retention. In fact, several studies have demonstrated conditions in which an expanding schedule may be equivalent or inferior to a uniform one (cf. Karpicke & Roediger, 2007, 2010). One interesting point this work has highlighted is the importance of pinpointing the optimal amount of spacing for the very first interval. If the first interval is too long then retrieval will fail, and the subsequent difficulty schedule matters little as learners are unlikely to recover the item (assuming there is no feedback during practice). If the first interval is too short then retrieval will succeed but may provide so little benefit that the expanding schedule is inferior to a uniform lag schedule with an equivalent amount of overall spacing that offers a longer initial interval.

Thus, implementing an optimally difficult practice plan hinges on determining the optimal initial amount of spacing, for a given learner studying a given item. Unfortunately, this is impossible. Empirically establishing whether a learner can retrieve an item after a particular lag requires testing that item. If the participant successfully retrieves the item, it is consequently strengthened, thereby invalidating the lag estimation. Conversely, if the participant fails to retrieve the item, the experimenter knows only that the lag was too long. This is an inherent limitation in the use of spacing to optimize learning, and

may explain why results based on expanding practice schedules have been mixed (Balota et al., 2007): the initial interval used in a particular study will vary in how close it is to what would be optimal for a given learner and item, and this proximity will influence the effectiveness of any subsequent lag schedule. Furthermore, there are practical limitations on the total time available during studying as well as difficulties that arise from the complexity of co-scheduling massive numbers of items at different stages of learning. However, even in the face of these limitations, Pavlik and Anderson (2008) developed an impressive model that makes ongoing estimates of optimal spacing schedules for practice of a large number of items, based on history of practice during a study session (for a given learner and item), and parameters estimated from prior data. Although the optimality of interval estimates must logically be limited for the earliest trials for each item, the model nevertheless produced effective overall schedules.

Cue informativeness as an alternative manipulation of difficulty

An alternative way of manipulating difficulty is to vary the amount of information provided by the retrieval cue (e.g., the number of letters of a to-be-remembered word; Benjamin, 2005; Carpenter & DeLosh, 2006; Carroll & Nelson, 1993). In spacing paradigms, difficulty increases with lag; in cue informativeness paradigms, difficulty increases with the poverty of the cue. Varying cue informativeness across practice trials, or *fading cues*, in order to enhance learning, is an idea that dates back to Skinner (1958) and includes variants such as the *vanishing cues* procedure (Glisky, Schacter, & Tulving, 1986), which we address further in the general discussion.

These approaches to the strategic regulation of difficulty are conceptually similar to common procedures of *computerized adaptive testing* (cf. Weiss & Kingsbury, 1984), which attempt to pinpoint a person's general ability or achievement in some domain by dynamically administering items of varying normative difficulty according to the person's successive responses. In contrast to such testing for the sake of assessment, our goal here was to exploit the powerful effects of testing as a learning event (cf. Roediger & Karpicke, 2006) in order to optimize learning of specific materials.

In the present study, we adopted this difficulty manipulation and examined two basic schedules for varying cue informativeness: *accumulating cues* (AC), and *diminishing cues* (DC). The effectiveness of both schedules was compared against a *study-only* control condition in which the entire target was presented on all trials.

In the DC condition, the informativeness of cues decreases across trials: initially easy practice becomes more difficult. In this way, DC is analogous to expanding retrieval practice; it minimizes the probability of retrieval failure while also ensuring that later trials are more difficult than early trials. We therefore predicted it would produce the greatest final recall performance of all the conditions.

In the AC condition, the informativeness of cues increases across trials: initially difficult practice becomes

more easy. Although AC is similar to the oft-dismissed condition of contracting retrieval practice, it differs in an important way: it offers learners a chance for recovery from retrieval failure on preceding trials. Furthermore, AC also offers greater opportunity than DC for more *potent* retrievals; the less well-learned an item is, and the more effortful its retrieval is, the greater the boost it will receive upon successful retrieval (Bjork & Bjork, 1992; Carpenter & DeLosh, 2006). This reasoning is also consistent with work by Gardiner, Smith, Richardson, Burrows, and Williams (1985), who found that the final-free-recall benefits of generating versus reading a word increased as a linear function of the number of letters omitted. AC, which presents the most difficult trials earliest in practice, is better suited than DC to allow such powerful retrievals. In other words, AC is more likely to hit the learner's sweet spot at least once. DC, by presenting the easiest trials earliest, may sell learners short of the opportunity to make high-yield effortful retrievals, while at the same time maximizing the number of overall successes.

In short, accumulating cues should promote potent retrieval at the cost of success; diminishing cues should promote successful retrieval at the cost of potency. Diminishing cues, like expanding retrieval practice, are expected to lead to an overall advantage in final recall.

Experiment 1

Experiment 1 was designed with two goals: to evaluate the prediction that diminishing cues would promote superior long-term retention; and to explore the trade-off between the potency and success of retrieval. An initial presentation of English–Iñupiaq word pairs was followed by practice trials (with no feedback) across which letters of the target word were either accumulated, diminished, or always present. A final test of cued recall measured the learning benefits of the three practice conditions.

Method

Participants

Eighty undergraduates participated for course credit.

Materials

Materials were 12 English–Iñupiaq word pairs (e.g., *dust–apyuq*), all nouns. The Iñupiaq words were all 5 letters long, and the English words varied in length from 3 to 7 letters. See the Appendix for the complete list.

Design

The experiment used a within-subjects design with one independent variable, practice condition, which had three levels: accumulating cues (AC), diminishing cues (DC), and study-only. The dependent measure was performance on a final cued recall test.

Procedure

Participants were run individually on computers programmed with Matlab using the Psychophysics Toolbox

extensions (Brainard, 1997). The procedure consisted of three phases: initial presentation, practice, and final test.

Initial presentation phase

Participants studied all 12 English–Iñupiaq word pairs, three times each, for 4 s per presentation, with no interstimulus intervals. The 12 word pairs were separated into three groups of four words each. Each group was rotated through the three practice conditions every three participants. For each participant, a random presentation order was generated for the initial presentation of the word pairs, and this order was repeated three times. Initial presentation was followed by a 1.5 min distractor task consisting of arithmetic problems and judgments about which of two circles was a darker shade of gray.

Practice phase

Participants completed six practice trials for each word pair. Each practice trial displayed an English word along with 0–5 (all) letters of the corresponding Iñupiaq word. Participants were instructed to type the full Iñupiaq word on each trial (even if all 5 letters of that word were provided), or to type a question mark if they did not know the word. There was no time limit and no feedback was given.

The three practice conditions are illustrated in Table 1. For word pairs in the AC condition, the first practice trial showed five underscores for the Iñupiaq word (e.g., *dust - _ _ _ _ _*), and each subsequent trial incrementally added one letter, until the sixth trial finally showed all letters. The order in which letters were added was determined randomly, with each added letter persisting on subsequent trials. Letters were always shown in their correct position, and underscores were spaced so as to clearly show the number of missing letters. The DC condition was the reverse of the AC condition: the first practice trial showed the entire Iñupiaq word, and each subsequent trial removed one letter. The study-only condition showed the entire Iñupiaq word on each trial, and instructed participants to type that word.

Practice trials were presented according to one of two fixed randomized orders (one of which was the reverse of the other). These orders were structured such that participants completed all practice trials for half of the pairs (an equal number from each condition) before moving on to practice trials for the other half of the pairs. Within these halves of practice, each pair received its first trial before any other pair received its second, and so on with later trials. The mean lag between successive practice trials for a given pair was 5 intervening trials (range: 3–7). The practice phase was followed by 10 min of the same distractor task described above.

Final test phase

All word pairs were tested using cued recall in an order that was randomized for each participant. Each test trial showed an entire English word along with five underscores (e.g., *dust - _ _ _ _ _*) and participants were instructed to type the full Iñupiaq word, or to type a question mark if they did not know the word. There was no time limit and no feedback was given.

Table 1
Examples of practice trials for the three practice conditions.

Practice trial	English word displayed	Iñupiaq word displayed		
		Accumulating cues (AC)	Diminishing cues (DC)	Study-only
1	dust -	_ _ _ _ _	a p y u q	a p y u q
2	dust -	_ _ _ u _	a p _ u q	a p y u q
3	dust -	_ _ y u _	_ p _ u q	a p y u q
4	dust -	a _ y u _	_ p _ u _	a p y u q
5	dust -	a _ y u q	_ p _ _ _	a p y u q
6	dust -	a p y u q	_ _ _ _ _	a p y u q
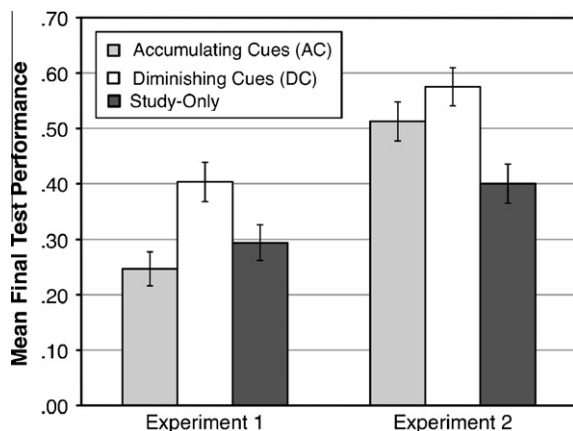
Fig. 1. Mean final recall performance as a function of practice condition and experiment. Error bars represent the standard error of the mean of each condition.

Results and discussion

An alpha level of .05 was used for all tests of statistical significance. No corrections for multiple comparisons were made because comparisons were few and were pre-planned. Effect sizes for comparisons of means are reported as Cohen's d calculated using pooled standard deviations (Olejnik & Algina, 2000, Box 1 Option B). Effect sizes for ANOVAs are reported as $\hat{\omega}^2$ (for one-way) or $\hat{\omega}^2_{partial}$, calculated using the formulae provided by Maxwell and Delaney (2004, pp. 547, 598). All within-subjects and mixed ANOVAs used uncorrected degrees of freedom, as Mauchly's test indicated that the assumption of sphericity was met in all instances.

Effectiveness of practice on final recall performance

Final recall performance as a function of practice condition is shown in Fig. 1 (left side), and in Table 2 along with practice performance. A one-way ANOVA revealed a reliable effect of practice condition, $F(2, 158) = 9.02$, $MS_e = .057$, $p < .001$, $\hat{\omega}^2 = .043$. DC recall was reliably greater than study-only recall, $t(79) = 2.99$, $p = .004$, $d = .37$. There was no reliable difference between AC recall and study-only recall, $t(79) = 1.28$, $p = .205$, $d = .17$. Of key interest, DC recall was reliably greater than AC recall, $t(79) = 3.92$, $p = .001$, $d = .53$. This result supports the hypothesis that,

much like expanding intervals in a spacing paradigm, the provision of diminishing retrieval cues leads to superior retention.

Differential potency of successful retrieval

Fig. 2 illustrates the benefit to final recall performance yielded by different numbers of successful practice retrievals for the AC and DC conditions. For each participant we calculated mean final recall performance as a function of practice condition and number of successful practice retrievals during practice for a given item (0–5; horizontal axis).[3] The center of each bubble corresponds to the mean of these participant means (vertical axis). The diameter of each bubble corresponds to the number of items in that cell, averaged across participants (range: 0.04–1.96).

There are two important patterns to note in Fig. 2. First, the large size of the rightmost white bubble indicates that DC practice yielded many instances of five successful practice retrievals for an item, while the large size of the leftmost gray bubble indicates that AC practice yielded many instances of zero practice retrievals for an item. A within-subject comparison of *total* number of successful practice retrievals for AC versus DC confirmed that successful practice retrieval was indeed reliably less frequent in AC than in DC ($M_{AC} = 5.8$, $SD_{AC} = 4.3$, $M_{DC} = 11.5$, $SD_{DC} = 6.0$), $t(79) = 9.34$, $p < .001$, $d = 1.09$. Second, the consistently higher vertical positions of the gray bubbles (AC) over the white bubbles (DC) indicate that *successful* AC practice retrievals yielded greater final test performance than successful DC practice retrievals. A comparison of the mean of participant means across number of successful practice retrievals ($n = 6$, as in Fig. 2) for AC versus DC showed that successful practice retrieval in AC was marginally more effective at enhancing final recall performance than was successful practice retrieval in DC ($M_{AC} = .45$, $SD_{AC} = .29$, $M_{DC} = .27$, $SD_{DC} = .26$), $t(5) = 2.20$, $p = .079$, $d = 0.66$.

These results illustrate the delicate balance between increased difficulty and increased probability of retrieval success (cf. Pyc & Rawson, 2009). The DC schedule provides a way of increasing the number of successful practice retrievals, but it also attenuates the benefits of these successes because earlier trials may be too easy. The AC schedule provides a way of maximizing the benefit of an initial successful practice retrieval because the first successful retrieval will necessarily be near the maximum level of difficulty at which a participant is capable of

---

[3] As shown in Table 1, both AC and DC practice conditions include one trial that amounts to a copy task. We exclude these trials from analysis.

Table 2
Means and standard deviations of performance across practice trials and on final test.

Practice condition	Practice trial						Final test
	1	2	3	4	5	6	
Experiment 1							
Accumulating cues	.12 (.32)	.12 (.32)	.18 (.38)	.25 (.43)	.45 (.50)	.95 (.22)	.25 (.27)
Diminishing cues	.89 (.31)	.69 (.46)	.63 (.48)	.54 (.50)	.51 (.50)	.51 (.50)	.40 (.32)
Study-only	.91 (.29)	.96 (.20)	.97 (.17)	.98 (.15)	.97 (.18)	.96 (.20)	.29 (.46)
Experiment 2							
Accumulating cues	.14 (.35)	.50 (.50)	.73 (.45)	.88 (.33)	.94 (.24)	.98 (.16)	.51 (31)
Diminishing cues	.93 (.26)	.78 (.42)	.81 (.39)	.78 (.41)	.78 (.42)	.83 (.37)	.58 (.31)
Study-only	.94 (.24)	.96 (.20)	.98 (.15)	.96 (.20)	.98 (.14)	.98 (.16)	.40 (.49)
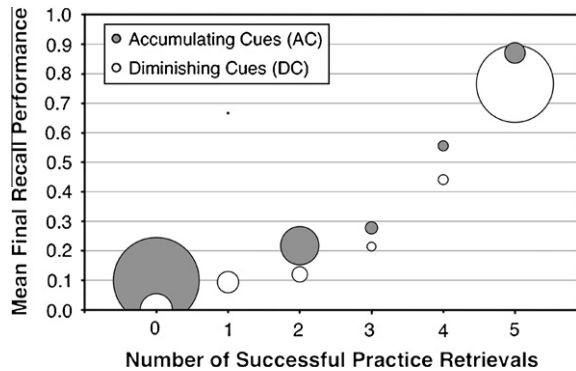
Fig. 2. Final recall performance (mean of participant means) as a function of practice condition (AC versus DC) and number of successful practice retrievals during practice for a given item (1–6). Bubble diameter represents mean number of items in each category, across participants (range: 0.15–1.95; Experiment 1).
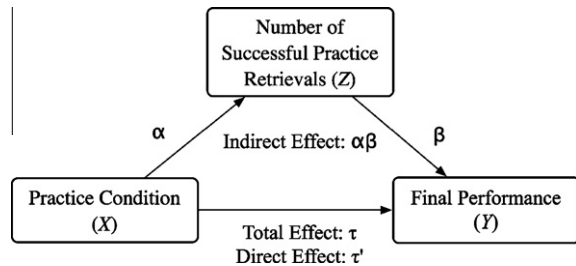
Fig. 3. Three-variable mediation model.

successfully retrieving that item. But this maximized benefit comes at the cost of diminishing benefits of further successes (because subsequent trials get easier) and at the cost of fewer overall successes than in the DC schedule.

The results considered so far suggest that the advantage of DC is due exclusively to the greater number of successful practice retrievals it promotes. That is, the greater probability of success under DC appears to outweigh the greater potency afforded by retrievals in AC. To evaluate this hypothesis more directly, we performed a mediation analysis based on the three-variable model illustrated in Fig. 3, with number of successful practice retrievals as a mediator between practice condition and final recall performance (cf. MacKinnon, Fairchild, & Fritz, 2007; MacKinnon, Krull,

Table 3
Estimated parameters of three-variable mediation model (see Fig. 3).

Effect	Parameter	Estimate	SE	Z	p
Experiment 1					
	α	1.50	0.14	10.94	<.001
	β	0.92	0.08	11.70	<.001
Total effect	τ	0.96	0.20	4.87	<.001
Direct effect	τ'	−0.55	0.27	−2.02	.044
Indirect effect	$\alpha\beta$	1.39	0.17	7.98	<.001
Experiment 2					
	α	0.29	0.09	3.44	.001
	β	0.82	0.09	9.24	<.001
Total effect	τ	0.34	0.18	1.93	.053
Direct effect	τ'	0.12	0.19	0.63	.528
Indirect effect	$\alpha\beta$	0.24	0.08	3.21	.001

& Lockwood, 2000). We estimated the parameters in Fig. 3 (, , τ, and τ') with linear mixed models with random intercepts for participants and items, using the following equations:

$$\text{logit}(P(Y_{ijk})) = \gamma_{jk} + \tau X_{ijk} \tag{1}$$

$$\text{logit}(P(Y_{ijk})) = \gamma_{jk} + \tau' X_{ijk} + \beta Z_{ijk} \tag{2}$$

$$Z_{ijk} = \gamma_{jk} + \alpha X_{ijk} + \varepsilon_{ijk} \tag{3}$$

where $\gamma_{jk} = \gamma_{00} + u_{j0} + v_{0k}$, i is the trial index, j is the participant index, k is the item index, γ_{00} is the overall intercept, u_{j0} is the deviation of a participant's intercept from the overall intercept, v_{0k} is the deviation of an item's intercept from the overall intercept, $_{ijk}$ is the error term, X is the practice condition (AC = 0, DC = 1, Study-only condition excluded from analysis), Y is final test performance (0, 1), and Z is the number of successful practice retrievals (0–5).

Table 3 shows the estimated parameter values. Using the Sobel (1982) method of estimating the indirect effect, and the Aroian (1944) method of estimating the standard error for this effect, we found that the number of successful practice retrievals indeed reliably mediated the effect of practice condition on final test performance (indirect effect = $\alpha\beta$) such that DC practice yielded superior final test performance overall (total effect = τ), but AC practice yielded superior final test performance when the number of successful practice retrievals was accounted for (direct effect = τ'). The opposite sign of the coefficients for the total and direct effects indicates an inconsistent mediation (MacKinnon et al., 2000) and more specifically, that the

number of successful retrievals is acting as a distorter variable that changes the direction of relation between two other variables (MacKinnon, 2008, p. 7; Rosenberg, 1968, p. 102). These results suggest that successful practice retrievals were more potent under AC conditions than DC conditions, but that final test performance was superior for DC because it permitted sufficiently more successful retrieval attempts to offset their reduced potency. Put differently, AC practice offers you more bang for your buck, but less buck.

Experiment 2

In Experiment 1, recall in the AC condition actually provided an advantage relative to DC when final recall performance was conditionalized on the number of successful practice retrievals. It is therefore surprising that AC did not yield an advantage over the study-only control condition. In Experiment 2, we evaluated whether providing participants with correct-response feedback after each practice trial would decrease the costs of retrieval failure but maintain the benefits of retrieval success. If so, the AC condition should be superior to the study-only control, in which no opportunities for effortful retrieval were provided (i.e., participants merely copied the correct response). In addition, the advantage of DC over AC should be less than it was when no feedback was provided (as in Experiment 1); feedback provides a failsafe, enabling participants to more quickly recover from early failed retrieval attempts (Kornell, Hays, & Bjork, 2009) and thus capitalize on more of the potent early AC trials.

Method

Participants, materials, and design
Eighty undergraduates participated for course credit, and the materials and design were identical to those employed in Experiment 1.

Procedure
The procedure was identical to that used in Experiment 1, with the following addition: on each practice trial, after a participant's response was made, the complete correct Iñupiaq word appeared on the screen for 4 s. Such feedback was supplied in all three practice conditions (including study-only) and was not contingent on participants' responses.

Results and discussion

Effectiveness of practice on final recall performance
Final recall performance as a function of practice condition is shown in Fig. 1 (right side), and in Table 2 along with practice performance. A one-way ANOVA revealed a reliable effect of practice condition, $F(2, 158) = 12.77$, $MS_e = .049$, $p < .001$, $\hat{\omega}^2 = .047$. DC recall was again reliably greater than study-only recall, $t(79) = 4.90$, $p < .001$, $d = .56$. As predicted—and unlike in Experiment 1—AC recall was also reliably greater than study-only recall, $t(79) = 3.22$, $p = .002$, $d = .36$. This effect was in contrast to the pattern
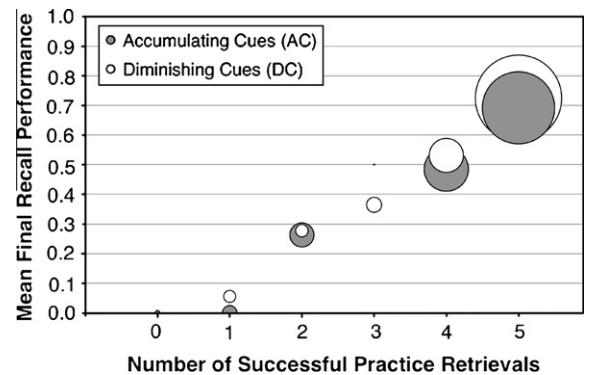

Fig. 4. Final recall performance (mean of participant means) as a function of practice condition (AC versus DC) and number of successful practice retrievals during practice for a given item (1–6). Bubble diameter represents mean number of items in each category, across participants (range: 0.00–2.15; Experiment 2).

in Experiment 1, as confirmed by a two-way ANOVA (Experiment 1 versus 2; AC versus. study-only) which revealed a reliable interaction, $F(1, 158) = 9.88$, $MS_e = .051$, $p = .002$, $\hat{\omega}^2_{partial} = .016$. This result supports our prediction about the effects of feedback in ameliorating some of the negative consequences of accumulating cues.

In Experiment 2, DC recall was only marginally greater than AC recall, $t(79) = 1.81$, $p = .075$, $d = .20$. To assess whether the advantage of DC over AC practice is reduced with feedback, we conducted a two-way ANOVA (Experiment 1 versus 2; AC versus DC), which revealed a marginally reliable interaction $(F(1, 158) = 3.15$, $MS_e = .056$, $p = .078$, $\hat{\omega}^2_{partial} = .004)$, providing some support for our prediction that feedback would increase the efficacy of the AC schedule relative to the DC schedule.

Differential potency of successful retrieval
Fig. 4 illustrates the benefit to final recall performance yielded by different numbers of successful practice retrievals for the AC and DC conditions in Experiment 2. Successful practice retrieval was again reliably less frequent in AC than in DC ($M_{AC} = 15.0$, $SD_{AC} = 4.3$, $M_{DC} = 16.2$, $SD_{DC} = 3.7$), $t(79) = 3.14$, $p = .002$, $d = 0.29$, but this disparity was reliably smaller than it had been in Experiment 1, $t(158) = -6.36$, $p < .001$, $d = -1.01$. As one would expect from the effects of feedback, and unlike the results in Experiment 1, successful practice retrieval was not more potent in the AC than in the DC condition ($M_{AC} = .32$, $SD_{AC} = .26$, $M_{DC} = .39$, $SD_{DC} = .23$), $t(5) = 0.09$, $p = .94$, $d = .28$. Thus, it appears that feedback, by making all practice trials less difficult, indeed increased the number of successful AC retrievals, but also attenuated the relative potency of these retrievals.

As with Experiment 1, we next turn to a mediation analysis in order to elucidate the relationship between practice condition, number of successful practice retrievals, and final recall performance (see Fig. 3 and Table 3). We again found that the number of successful practice retrievals reliably mediated the effect of practice condition on final test performance (indirect effect =) such that DC practice yielded marginally superior final test performance overall

(total effect = τ), but the effect of practice condition was not reliable when the number of successful practice retrievals was accounted for (direct effect = τ'). Thus, although number of successful practice retrievals still mediated the effect of practice condition on final test performance, it did not reverse that effect as in Experiment 1. This was likely due to the presence of feedback in Experiment 2 both diluting the potency and increasing the prevalence of successful AC practice retrievals.

General discussion

A major challenge in implementing desirable difficulties in learning is effectively calibrating that difficulty across trials. An optimal design strategy would account for individual variations in learners and in materials, such that each retrieval is neither so difficult that it fails nor so easy that it provides negligible benefit. Toward this end, we have demonstrated here a manipulation of difficulty that is a useful alternative to spacing: cue informativeness. A schedule of accumulating cues (increasing cue informativeness across practice trials) automatically locates the highest level of difficulty at which a certain learner can first retrieve a certain item. Furthermore, unlike difficulty schedules based on spacing (e.g., expanding versus contracting retrieval practice), there are potential benefits to both accumulating and diminishing cues. Accumulating cues practice promotes effortful retrieval and its consequent mnemonic benefits, at the cost of likely success; diminishing cues practice promotes successful retrieval, at the cost of those benefits.

An additional possible benefit afforded by AC practice is suggested by recent work by Kornell et al. (2009; see also Izawa, 1970; Richland, Kornell, & Kao, 2009). They found that a failed retrieval attempt can enhance the benefits of subsequent opportunities to study the item in question. Thus, presenting a learner with a trial that is too difficult (as may be the case for early AC trials) is not necessarily a waste of time and effort, as long as it is remedied by further information to ultimately ensure a subsequent successful retrieval, either in the form of greater cue informativeness on later trials or in the form of feedback. Such an effect may have contributed to the AC potency advantage we have observed here, but confirming the presence of this effect requires further study. Further study should also investigate performance at longer retention intervals (e.g., days rather than minutes), as any benefits of AC practice may become more apparent at a longer delay (cf. Karpicke & Roediger, 2007).

Finally, it is worth noting that the discrete nature of accumulating/diminishing individual letters offers a coarser adjustment of difficulty than would spacing, which can be varied on a much finer scale. However this does not eliminate the advantage that cue informativeness can better be used to identify appropriate levels of difficulty. Furthermore, higher-precision manipulations of cue informativeness are possible, such as varying the focus or occlusion of picture stimuli for a recognition task.

Individual differences in relative benefits of practice conditions

Riley and Heaton (2000) explored the benefits of a similar paradigm for patients with a history of head injuries and how those benefits varied with an individual's current performance. They found that decreasing assistance (analogous to diminishing cues) was more effective than increasing assistance (analogous to accumulating cues) for patients with poorer memory, but the opposite was true for those with better memory. In a similar vein, McDaniel, Hines, and Guynn (2002) found that studying text passages with some letters deleted improved recall of propositions for skilled readers, but impaired recall for unskilled readers (compared to a control condition with no deleted letters).

We reasoned that the relative benefits of practice condition in our paradigm should similarly vary as a function of individual differences in initial learning, considering that retrieval difficulties are beneficial only to the extent that they can be overcome. Participants who did not learn the material well by the end of the initial presentation phase likely floundered with AC practice and benefited from the early assistance provided by DC. Conversely, participants who learned the material well enough to meet the early challenge of AC should have shown a lesser advantage of DC over AC. We investigated this possibility by calculating correlations between mean performance on the first practice trials of the AC condition (our best available indicator of initial learning) and the difference in final test performance for DC versus AC. These correlations were reliably negative for Experiment 1, $r = -.32$, $t(78) = -3.03$, $p = .003$, and for Experiment 2, $r = -.24$, $t(78) = -2.11$, $p = .038$. These results suggest that individual differences in initial learning moderated the extent to which participants were able to capitalize on the opportunity for potent retrievals provided by AC practice.

Related prior work using cue informativeness

Although learners may themselves employ some desirable difficulties during self-directed learning, the management and optimization of complex difficulty schedules (whether based on spacing, cue informativeness or some other manipulation) are likely best implemented through well-designed learning environments, which can themselves serve to improve learners' metacognition as well as their learning (Finley, Tullis, & Benjamin, 2010). A framework from educational research that encompasses this approach is that of scaffolding: shaping instruction to assist a learner to perform beyond the level s/he is currently independently capable, with the goal of ultimately removing this assistance (cf. Lajoie, 2005; Linn, 1995; Pea, 2004; Wood, Bruner, & Ross, 1976). The research we have described here can be considered an instance of scaffolding, as can other studies that have used cue informativeness, which we will now review.

A little-known set of efforts have been made to apply a strategy of progressive letter deletion (analogous to DC) to

memorization of more complex verbal materials such as poetry and prose (Boyd, 1989; Carlson, 1981; Schilmoeller, Schilmoeller, & King, 1982). Although the results of this work have been inconclusive, such efforts highlight additional important factors in the use of cue informativeness procedures, such as the rate of letter removal, the pattern of letter removal (e.g., whether the same letters removed on one trial will remain absent on subsequent trials), and truncation of the practice schedule (i.e., halting practice before all letters are removed). Furthermore, McDaniel, Einstein, Dunay, and Cobb (1986) propose that the utility of a difficulty manipulation depends on the kind of processing it promotes *and* the kind of processing naturally invited by the learning material. For example, they claim that progressively removing letters from a text across study trials promotes individual-item processing (as opposed to relational processing) and will thus most benefit texts that did not already encourage such processing (e.g., narratives). Future research should therefore also manipulate the nature of the to-be-learned material in an AC/DC paradigm.

More relevant prior work has been done in the clinical context of cognitive rehabilitation for individuals with memory impairment or developmental disabilities (cf. Wilson, Herbert, & Shiel, 2003, pp. 50–68). This work may be found under terms such as: *increasing* versus *decreasing assistance, fading cues, prompt hierarchy, hierarchy of cues,* and *vanishing cues*. Whereas our goal here has been to help learners retrieve all material at the hardest level possible for them, these related paradigms have placed more emphasis on helping learners avoid errors of commission (i.e., false recalls). Such errors are particularly problematic for memory-impaired populations, because the errors can tend to propagate in the absence of intact explicit learning processes to correct them (Baddeley & Wilson, 1994). Thus, to better capitalize on intact implicit learning abilities in such individuals (Baddeley, 1992), researchers have developed a method of *errorless learning* in which participants merely copy the correct answer to a stem completion task instead of attempting to guess or retrieve it and then receiving feedback (cf. Kessels & de Haan, 2003). One disadvantage to this approach is that it sells short individuals with any intact explicit abilities because it does not offer any opportunities for retrieval.

Glisky et al. (1986) proposed a method of vanishing cues as a way of still reducing errors yet also allowing for effortful retrieval to the extent that the individual is capable. In this method, cue letters accumulate within a trial (beginning at zero cues for the first trial), then diminish across subsequent trials (with each trial beginning at one fewer than the number of cues needed for success on the previous trial). This method has shown mixed benefit for memory impaired patients and healthy control groups when compared to a control condition of standard anticipation trials (cf. Glisky et al., 1986; Hunkin & Parkin, 1995; Kessels & de Haan, 2003; Riley & Heaton, 2000; Riley, Sotiriou, & Jaspal, 2004), possibly due to perseverance of incorrectly guessed answers, particularly for patients with impaired explicit memory.

There are a few key differences between such prior work and the present study. First, prior work has often included no initial presentation phase. Second, feedback has almost always been given at the end of each trial. Finally, the difficulty schedules used have typically been implemented contingent on participants' performance. For example, cues may be added (typically within a trial) only after a certain number of incorrect responses or a failure to respond within a certain amount of time, and cues may be removed (typically across trials) only after a performance criterion has been met. These factors, while sensible in clinical contexts, make it difficult to determine the specific causes of any enhanced performance.

The present research analyzes the recall performance contributions of the two basic types of difficulty schedules in their purest form (accumulating cues and diminishing cues). This comparison is important to better inform more advanced hybrid approaches that: (a) combine elements of both accumulating and diminishing cues, and (b) adapt according to learners' responses. For example, accumulating cues within a trial and diminishing cues across trials could constitute a self-calibrating desirable difficulty that better achieves what cannot be achieved by manipulating spacing alone: pinpointing the optimal level of difficulty on the first practice trial for a given participant and item, and also ensuring subsequent retrieval that is both successful and maximally effortful. Scheduling of repeated practice (i.e., spacing) would still have a role to play in such a learning plan, but the need to precisely determine the optimal amount of spacing for any trial would be greatly alleviated.

Summary

Any successful learning regimen must calibrate the difficulty of a retrieval opportunity to the current level of learning. If retrieval is too easy, the resulting benefits to learning are minimal; if it is too difficult, the probability of successful retrieval is undesirably low. The current research demonstrates that the benefits of expanded retrieval are not limited to paradigms in which scheduling is manipulated; rather, any method by which the difficulty of retrieval can be systematically varied (e.g., accumulating and diminishing cue informativeness) is a potential candidate for harnessing the value of desirable difficulties and helping to optimize learning.

Acknowledgments

Appendix

Materials	
English word	Iñupiaq word
boot	kamik
cliff	ikpik
duck	mitiq
dust	apyuq
heaven	qixak
kidney	taqtu
mark	aglak
rainbow	nigaq
sled	uniat
snail	uvixu
tea	saiyu
thread	ivalu

References

Aroian, L. A. (1944). The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics, 18*, 265–271.

Baddeley, A. D. (1992). Implicit memory and errorless learning: A link between cognitive theory and neuropsychological rehabilitation? In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (2nd ed., pp. 309–314). New York, NY, US: Guilford Press.

Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia, 32*, 53–68.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*, 296–308.

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 83–105). New York, NY, US: Psychology Press.

Benjamin, A. S. (2005). Response speeding mediates the contribution of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review, 12*, 874–879.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Boyd, J. H. (1989). An analysis of the effects of progressive letter deletion on free recall of prose: encoding variability, differential processing, or fading of prompts? Unpublished master's thesis, University of Notre Dame, Indiana.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Carlson, N. R. (1981). LEARNIT: A memorization-aid program. *On Computing, 3*, 72–78.

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

Carroll, M. C., & Nelson, T. O. (1993). Failure to obtain a generation effect during naturalistic learning. *Memory & Cognition, 21*, 361–366.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*, 1095–1102.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43*, 627–634.

Finley, J. R., Tullis, J. G., & Benjamin, A. S. (2010). Metacognitive control of learning and remembering. In M. S. Khine & I. Saleh (Eds.), *New science of learning: cognition, computers and collaboration in education*. Springer.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1*, 213–216.

Gardiner, J. M., Smith, H. E. C., Richardson, C. J., Burrows, M. V., & Williams, S. D. (1985). The generation effect: Continuity between generating and reading. *American Journal of Psychology, 98*, 373–378.

Glisky, E. L., Schacter, D. L., & Tulving, E. (1986). Learning and retention of computer-related vocabulary in memory-impaired patients: Method of vanishing cues. *Journal of Clinical and Experimental Neuropsychology, 8*, 292–312.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*, 392–399.

Greene, R. L. (2008). Repetition and spacing effects. In H. L. Roediger, III (Ed.). *Learning and memory: A comprehensive reference. Cognitive psychology of memory* (Vol. 2, pp. 65–78). Oxford: Elsevier.

Hunkin, N. M., & Parkin, A. J. (1995). The method of vanishing cues: An evaluation of its effectiveness in teaching memory-impaired individuals. *Neuropsychologia, 33*, 1255–1279.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*, 340–344.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704–719.

Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38*, 116–124.

Kessels, R. P. C., & de Haan, E. H. F. (2003). Implicit learning in memory rehabilitation: A meta-analysis on errorless learning and vanishing cues methods. *Journal of Clinical and Experimental Neuropsychology, 25*, 805–814.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989–998.

Lajoie, S. P. (2005). Extending the scaffolding metaphor. *Instructional Science, 33*, 541–557.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: The scaffolded knowledge integration framework. *Journal of Science Education and Technology, 4*, 103–126.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593–614.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science, 1*, 173–181.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language, 25*, 645–656.

McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language, 46*, 544–561.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241–286.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*, 101–117.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13*, 423–451.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437–447.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*, 243–257.

Riley, G. A., & Heaton, S. (2000). Guidelines for the selection of a method of fading cues. *Neuropsychological Rehabilitation, 10*, 133–149.

Riley, G. A., Sotiriou, D., & Jaspal, S. (2004). Which is more effective in promoting implicit and explicit memory: The method of vanishing cues or errorless learning without fading? *Neuropsychological Rehabilitation, 14*, 257–283.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.

Schilmoeller, G. L., Schilmoeller, K. J., & King, D. R. (1982). An empirical analysis of "LEARNIT," a microcomputer program for facilitating memorization. In J. Smith & G. Schuster Moun (Eds.), *Proceedings of National Educational Computing Conference, 1982* (pp. 306–310). Columbia, MO: University of Missouri.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G. E. Stelmach & J.

Requin (Eds.), *Tutorials in motor neuroscience* (pp. 59–75). Dordrecht: Kluwer.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179–187.

Skinner, B. F. (1958). Teaching machines. *Science, 128*, 969–977.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38*, 244–253.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*, 361–375.

Wilson, B. A., Herbert, C. M., & Shiel, A. (2003). *Behavioural approaches in neuropsychological rehabilitation: Optimising rehabilitation procedures*. New York, NY: Psychology Press.

Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychiatry and Psychology, 17*, 89–100.