Effects of criterion level on associative memory: Evidence for associative asymmetry

Kalif E. Vaughn *, Katherine A. Rawson

Kent State University, United States

A R T I C L E  I N F O
Article history:
Received 27 June 2013
revision received 3 March 2014
Available online 21 May 2014

Keywords:
Human memory
Human learning
Criterion learning
Associative memory
Associative symmetry

A B S T R A C T
A wealth of research has established that retrieval practice promotes subsequent memory, particularly when the retrieval attempt is successful. Furthermore, the number of successful retrievals during practice (i.e., criterion level) dramatically influences final test performance. For example, Vaughn and Rawson (2011) had participants learn Lithuanian–English word pairs via test–restudy practice until they were correctly recalled. Despite retrieval practice always occurring in the forward direction during practice (A – ?), performance increased as a function of criterion level both on final forward (A – ?) and backward (? – B) cued recall tests. Importantly, the performance gain across criterion levels appeared asymmetric, as the gains were much larger in the forward versus backward cued recall direction. However, one potential explanation for the observed asymmetry in criterion level effects is that the materials strongly favored forward cued recall, as retrieving Lithuanian versus English is inherently more difficult for native English speakers. The present experiments utilized English–English pairs to more appropriately investigate whether the effects of criterion level on associative memory are symmetric or asymmetric. Across experiments, results from recall and recognition tests indicated that criterion level effects on associative memory are asymmetric. Advantages in forward versus backward cued recall could not be attributed to differences in cue memory or target memory, indicating differences in forward and backward associative memory.

Introduction
A wealth of research has demonstrated that testing enhances memory, often more than an equivalent amount of studying (e.g., for recent reviews, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & Butler, 2011). However, the effectiveness of testing depends on many factors, including the type of test format used and the retention interval between practice and final test (see Roediger & Karpicke, 2006 for a review). Another key factor influencing the effectiveness of testing is retrieval success. Testing enhances memory to a greater degree when the retrieval attempt is successful versus unsuccessful (e.g., Carrier & Pashler, 1992). Additionally, the effectiveness of successful retrieval practice is influenced by the number of successful retrievals during practice (i.e., criterion level). Recently, studies have investigated the effects of criterion level on subsequent memory (e.g., Pyc & Rawson, 2009; Rawson & Dunlosky, 2011; Vaughn & Rawson, 2011). Across these studies, the general finding is that memory is substantially enhanced with higher versus lower criterion levels (i.e., final test performance increases as the number of correct recalls during practice increases). In
the current research, we explored the nature of criterion level effects on subsequent memory, with particular focus on the gains in associative memory. To what extent is the effect of criterion level on associative memory symmetric (i.e., increasing criterion level produces similar gains in the forward and backward associations for cue–target word pairs) or asymmetric (i.e., increasing criterion level produces stronger gains in either the forward or backward association for cue–target word pairs)?

Although previous research has established that memory performance increases as criterion level increases, only one recent study has explored the nature of criterion level effects. Vaughn and Rawson (2011) had participants learn Lithuanian–English word pairs. After the word pairs had been presented for an initial study trial, participants completed practice cued recall tests until they correctly retrieved the word pairs a pre-assigned number of times (from one to five). Importantly, the practice cued-recall tests always occurred in the forward direction (i.e., participants were shown the Lithuanian word and retrieved the English translation). The final test phase assessed forward cued recall (the same test format used during practice) and backward cued recall (i.e., participants were shown the English word and retrieved the Lithuanian translation). Despite practicing only forward cued recall, performance increased as criterion level increased for both forward and backward cued recall. Although performance increased in both cued recall directions, performance gains were much greater in the forward versus backward direction (gains of 35% versus 12%, respectively). The discrepancy in performance gains between forward and backward cued recall raises an interesting theoretical question: Are the effects of criterion level on associative memory symmetric or asymmetric?

**Basis for predicting asymmetric effects of criterion level on associative memory**

Several theoretical accounts support the prediction that the effects of criterion level on associative memory are asymmetric (i.e., forward and backward links are encoded and strengthened independently and thus need not be equivalent). For example, according to the transfer appropriate processing framework (TAP; Morris, Bransford, & Franks, 1977), performance on a final test will be maximized when the initial tests and final tests require similar cognitive operations. If individuals practice in the forward direction, TAP would predict that performance on the final test will be maximized when individuals are again tested in the forward direction.

Although theories proposed to explain the benefits of retrieval practice in particular are silent on associative symmetry, reasonable extensions follow given the mechanisms proposed by these theories. For example, according to the elaborative retrieval hypothesis (ERH; e.g., Carpenter, 2009), attempting retrieval from a cue word activates related semantic information that can be encoded along with the retrieved target, which in turn can later serve as additional pathways to retrieve the target information. For instance, testing on the word pair morning – light (morning – ?) may activate associates of the cue word (e.g., breakfast, shower, commute, tired), all of which can be used later in the service of retrieving the target information (e.g., light). Presumably, the information activated during testing may be more strongly associated with the word serving as the cue (e.g., morning) versus the target (e.g., light), suggesting that the effects of criterion level on associative memory will be asymmetric (favoring forward versus backward associative memory).

Similarly, the mediator effectiveness hypothesis (MEH; e.g., Pyc & Rawson, 2010) states that retrieval practice benefits memory by promoting the use of more effective mediators (i.e., information linking cues to targets). For example, keyword mediators commonly involve words that share phonetic similarities to a particular cue in a cue–target word pair (e.g., wing for the Swahili–English word pair wingu – cloud). The mediator wing can then be used to link the mediator to the target word (e.g., birds have wings and fly in the clouds; wingu – wing – cloud). Importantly, these verbal mediators are likely to function asymmetrically (e.g., wingu elicits wing, but cloud does not elicit wing due to lack of phonetic similarity). If the benefits of testing occur due to more effective mediator use and these mediators tend to link cue–target information asymmetrically, the implication is that the effects of criterion level on associative memory will be asymmetric (again favoring forward versus backward associative memory).

For clarification purposes, although both ERH and MEH support the prediction that the effects of criterion level on associative memory will be asymmetric, the underlying mechanisms driving this prediction differ somewhat between the two accounts. Both accounts would attribute the difference in forward versus backward associative memory across criterion level to differences in the directionality of semantic information added to the nominal cue–target stimulus. However, the accounts differ regarding how and when this semantic information is added to the nominal cue–target stimulus. According to ERH, the additional semantic information is activated during a retrieval attempt, and reflects a more automatic and implicit process. In contrast, MEH assumes a more strategic process in which participants are consciously selecting keywords as part of a metacognitive control strategy to learn the words (which would likely occur during restudy that takes place after a retrieval attempt; for evidence, see Pyc & Rawson, 2012). Nonetheless, these accounts are not mutually exclusive, and either or both mechanisms could contribute to asymmetric associations.

Another relevant account comes from basic research on associative memory. According to the independent association hypothesis (IAH; Wolford, 1971), two elements in episodic memory are associated with two unidirectional links: one linking the two elements in a forward direction, and the other linking the two elements in a backward direction. Importantly, these separate unidirectional links can be enhanced independently of each other, such that learning in one direction can occur independently of learning in the other direction (e.g., greater learning in the forward versus backward direction).
Basis for predicting symmetric effects of criterion level on associative memory

In contrast, according to the associative symmetry hypothesis (ASH; Asch & Ebenholtz, 1962), two elements in episodic memory are associated with one bidirectional link (i.e., the same connection links two elements in a forward and backward direction). Importantly, the bidirectional link dictates that learning in one direction cannot occur independently of learning in the other direction (e.g., learning in the forward direction equally enhances memory in the backward direction). Thus, ASH would assume that the effects of criterion level on associative memory would be symmetric (in contrast to expectations that reasonably follow from the elaborative retrieval hypothesis, mediator effectiveness hypothesis, and IAH).

The expectation that the effects of criterion level on associative memory are symmetric is also supported by the available empirical evidence on associative symmetry. Given the voluminous literature pertaining to associative symmetry, an exhaustive review is beyond the scope of the present article. Fortunately, Kahana, (2002) recently reviewed the literature and concluded that the weight of the evidence heavily favors associative symmetry. This is not to say that there are no prior studies reporting evidence that appeared to demonstrate associative asymmetry (typically involving a demonstration of differences in forward versus backward cued recall). However, limitations of the methods used in these studies preclude clear interpretation of these outcomes. In particular, Kahana (2002) discusses two major methodological factors which need to be accounted for in order to evaluate associative symmetry. The first methodological issue concerns the extent to which the cue and target words have a similar number of pre-existing associates. Words with a higher number of pre-existing associates may produce more interference during cued recall, decreasing the chance of successfully retrieving the lesser associated word within the word pair (e.g., if the number of pre-existing associates is greater for a target word than for a cue word, the cue word would be more difficult to retrieve during backward cued recall due to more interference from other associates of the target).

There are several ways to minimize concerns over differences in pre-existing associations, including drawing items from the same class of stimuli (e.g., using noun–noun word pairs) and then counterbalancing assignment of the items to be cues and targets (e.g., A–B and B–A). Many prior studies violate one or both of these important prerequisites (e.g., Bartling & Thompson, 1977; Gallup & Wollen, 1968; Levy & Nevill, 1974; Lockhart, 1969) making it difficult to interpret their results with respect to associative symmetry. To foreshadow, in all of the current experiments, we minimized these concerns by using unrelated noun–noun word pairs and by counterbalancing assignment of words to serve as either the cue or target during practice (e.g., “queen–journal” and “journal–queen”).

The second methodological issue concerns accounting for differences in item accessibility, because items with higher accessibility are more easily retrieved during cued recall (e.g., higher levels of item accessibility for target versus cue words would benefit forward versus backward cued recall). The key here is that cued recall measures are not pure measures of associative memory. Performance on a cued recall test reflects associative memory and also reflects memory for the information being directly retrieved. For instance, performance on a final forward cued recall test reflects both associative memory and target memory. Likewise, performance on a final backward cued recall test reflects both associative memory and cue memory. Accordingly, associative symmetry cannot be adequately evaluated solely by comparing forward versus backward cued recall, because any differences may reflect differential accessibility of targets versus cues rather than associative asymmetry.

There are several ways to minimize concerns over differences in item accessibility, such as including measures to assess both cue and target accessibility and/or implementing a procedure to equate cue and target accessibility (e.g., additional exposure to cue or target words, depending on which items have lower accessibility). However, many prior studies did not measure cue or target accessibility and/or made no attempt to equate cue or target accessibility (e.g., Caplan, Glaholt, & McIntosh, 2006; Giurintano, 1972; Newman & Campbell, 1971; Wollen, 1968; Wollen & Allison, 1968). For example, many studies involved participants practicing in a forward direction (e.g., A – ?) followed by subsequently demonstrating superior forward versus backward cued recall (e.g., Gallup & Wollen, 1968; Giurintano, 1972; Lockhart, 1969). However, no measures of target or cue accessibility were included, and item accessibility differences likely arose due to differential strengthening of items during practice. Forward cued recall practice presumably enhances memory strength for the target word in addition to strengthening the cue–target association. If so, the demonstrated advantage of forward versus backward cued recall may have reflected greater target accessibility rather than associative asymmetry. To foreshadow, we minimized these concerns by including measures of cue and target accessibility in all experiments and by pre-exposing cues to equate cue and target accessibility in Experiment 4.

In sum, much of the prior evidence for associative asymmetry is not readily interpretable due to one or more of these methodological limitations. In contrast, numerous studies have shown evidence suggesting associative symmetry (e.g., Asch & Ebenholtz, 1962; Ekstrand, 1966; Mandler, Rabinowitz, & Simon, 1981; Murdock, 1962a, 1962b, 1965, 1966; Newman & Campbell, 1971). For example, Murdock (1962b) found similar forward and backward cued recall performance regardless of presentation rate (i.e., 1 s or 3 s of study time per paired associate), retention interval (recall started 0–10 s after the last displayed paired associate), or number of presentations (i.e., paired associates were presented either one, two, or three times during practice). More recent studies have also concluded in favor of associative symmetry (e.g., Carpenter, Pashler, & Vul, 2006; Jones & Pashler, 2007; Sommer, Rose, & Büchel, 2007; Yang et al., 2013).

With that said, Kahana (2002) notes that “Even if forward and backward recall are equivalent on average, this does not strictly imply symmetry” (p. 823). As outlined
above, forward and backward cued recall are not pure measures of associative memory but also reflect target memory or cue memory as well. Thus, the prior studies demonstrating similar levels of forward and backward recall do not definitively indicate associative symmetry in the absence of evidence that also indicates similar levels of target and cue memory. Furthermore, Kahana (2002) acknowledged that “It is possible that forward and backward associations reveal their asymmetries in other experimental situations” (p. 835), which may be the case when information is learned to varying criterion levels. Thus, the extent to which the effects of criterion level on associative memory are symmetric remains an open question.

Overview of current research

In sum, although the goal of the current work was not to competitively evaluate the various theoretical accounts described above, these accounts do support different expectations for how criterion level may affect associative memory. Furthermore, although the bulk of the empirical evidence favors associative symmetry (Kahana, 2002), some interpretive difficulties remain, and none of this research involved criterion level manipulations (the condition of primary interest here). Outcomes of the only criterion level study that included relevant measures (Vaughn & Rawson, 2011, described above) suggested that the effects of criterion level on associative memory are asymmetric. However, the study by Vaughn and Rawson (2011) was not originally designed to evaluate associative symmetry, and not surprisingly, the materials used violated the aforementioned methodological conditions necessary to evaluate associative symmetry. For example, the number of pre-existing associates differed between cue and target words, as Lithuanian versus English words have a lower number of pre-existing associates for native English speakers. Furthermore, performance on free recall and recognition measures suggested poorer item accessibility for cue words than for target words. Given these methodological limitations, the results of Vaughn and Rawson (2011) only provide suggestive evidence at best for associative asymmetry.

In the present experiments, we minimized concerns over potential differences in the number of pre-existing associates and pre-experimental item accessibility by using pairs of concrete English nouns (i.e., the same stimulus category) and by counterbalancing the assignment of the two words in each pair to serve as the cue versus the target (e.g., either “queen–journal” or “journal–queen”). Regarding item accessibility, we also estimated differences in overall item accessibility by examining the effects of criterion level on separate measures of target memory and cue memory in addition to forward and backward cued recall. During practice, participants learned English–English word pairs to criterion via forward cued recall practice (A → ?). Criterion level (i.e., the number of correct recalls) ranged from one to five. The practice phase for a given participant ended when all the items had reached their pre-assigned criterion level. The final test occurred one week later. During the final test, we administered forward and backward cued recall measures, as well as measures of cue memory and target memory (described further below).

To evaluate the extent to which criterion level effects on associative memory are symmetric, we evaluated patterns of forward and backward cued recall performance while also considering levels of target memory and cue memory. To revisit, cued recall measures are not pure measures of associative memory, and thus performance on a cued recall test reflects associative memory and also reflects memory for the information being directly retrieved. Accordingly, equivalent gains on the forward versus backward cued recall tests would not necessarily indicate symmetry unless the gains for target memory and cue memory were also equivalent. Likewise, greater performance on the forward versus backward cued recall tests would not necessarily indicate asymmetry if greater gains for target memory versus cue memory were also observed. Thus, evidence that the effects of criterion level on associative memory are symmetric would involve either (1) equivalent gains in forward and backward cued recall and equivalent gains in target memory and cue memory, or (2) greater gains on the forward versus backward cued recall test and greater gains for target memory versus for cue memory. Any other combination of outcomes would indicate that the effects of criterion learning on associative memory are asymmetric given that the differences in forward versus backward cued recall would not be attributable to differences in target versus cue memory. If differences in forward versus backward cued recall cannot be attributed to differences in target or cue memory, then these differences must reflect differences in associative memory.

Experiment 1

Method

Participants and design

One hundred three undergraduates participated for course credit. Criterion level during practice (1, 2, 3, 4, or 5 correct recalls) was a within-participant manipulation. Recall and recognition test formats were between-participant manipulations. First, participants were assigned to one of four recall groups via blocked randomization. Second, within each recall group, assignment to one of three recognition groups was counterbalanced across participants.

Materials

Materials included 50 unrelated English–English word pairs (e.g., queen–journal), divided into five lists of ten pairs. The cue–target pairs had zero forward and zero backward associative strength (Nelson, McEvoy, & Schreiber, 2004). Which word served as the cue versus the target was approximately counterbalanced across participants. List assignment to criterion level was approximately counterbalanced across participants.

Procedure

Participants first read instructions detailing the experiment. Instructions informed participants that during the
learning phase, they would study and take practice tests on English–English word pairs until they had correctly recalled them anywhere from 1 to 5 times. Participants were also informed that there would be a final test over the word pairs after a one-week delay; however, participants were not informed about the format of the final test (i.e., participants were simply told “you will take a final test of your memory for the items you learn today” but were given no specific information that they would be completing either a cued recall or free recall test followed by a recognition test). These instructions were consistent for each experiment reported within this manuscript. Word pairs were presented for initial study in random order one at a time via computer for 10 s each. The word on the left was always presented in blue font under the heading “cue word,” whereas the word on the right was always presented in red font under the heading “target word.” After all items were studied once, the practice phase began. During practice trials, participants were shown a cue word (always presented in blue font on the left) and had up to 8 s to type in the corresponding target. This test format represents forward cued recall (A – ?) and the only cued recall test format used during practice. After incorrect responses, the cue and target were restudied for 4 s. The item was then placed at the end of the list for another practice trial later. If a response was correct but the item had not yet reached its assigned criterion, it was placed at the end of the list for another practice trial later. Items were dropped from practice once they reached their pre-assigned criterion. Session 1 ended when all items reached criterion or after 75 min had elapsed. We excluded data for ten participants who did not reach at least 75% criterion for all items during the allotted time in Session 1.

In Session 2 seven days later, participants completed one of four recall tests, depending on group assignment. For the two cued recall groups, items were presented one at a time and presentation order was randomized anew for each participant. For the group of participants who completed forward cued recall, cues were presented in blue font on the left, and participants were instructed to type in the corresponding target (e.g., queen → ???). For the group of participants who completed backward cued recall, targets were presented in red font on the right and participants were instructed to type in the corresponding cue (e.g., ??? – journal). For the group who completed free recall of cues, participants were instructed to recall as many cue words as possible and were reminded that these were the words that had been presented in blue on the left side of the screen during Session 1. For the group who completed free recall of targets, participants were instructed to recall as many target words as possible and were reminded that these were the words that had been presented in red on the right. In all four groups, participants had an unlimited amount of time to type in their responses.

After the recall test, participants completed their assigned recognition test. The cue recognition test consisted of the 50 previously learned English cue words randomly mixed with 50 novel English words. The associative recognition test consisted of 50 English–English word pairs. Twenty-five of the English cues were paired correctly with their corresponding English targets, with five correct pairs from each criterion level. The remaining 25 English cues were paired incorrectly with an English target from another pair from the same criterion level. Participants were informed that all pairs contained words they had studied earlier and that they were to indicate whether the English cue was paired with its corresponding English target.1

Results

To revisit, evidence that the effects of criterion level on associative memory are symmetric would involve either (1) equivalent gains in forward and backward cued recall and equivalent gains in target memory and cue memory, or (2) greater gains on the forward versus backward cued recall test and greater gains for target memory versus for cue memory. First, consider forward versus backward cued recall performance (see Fig. 1). A 2 (type of cued recall test) × 5 (criterion level) mixed factor ANOVA revealed a main effect of criterion level, \( F(4,172) = 11.55, \text{MSE} = 1.80, p < .001, \eta_p^2 = .21 \), replicating previous research (e.g., Vaughn & Rawson, 2011). The overall performance difference between forward versus backward cued recall approached significance, \( F(1,43) = 3.60, \text{MSE} = 13.75, p = .065, \eta_p^2 = .08 \). Qualitatively, greater overall criterion level gains emerged for forward versus backward cued recall (22% versus 11%, as measured via the net performance gain from criterion level 1 to criterion level 5); however, this interaction did not reach significance, \( F(4,172) = 1.50, \text{MSE} = 1.80, p = .206, \eta_p^2 = .03 \). When including only the two extreme criterion levels (criterion level 1 and criterion level 5) in a 2 (type of cued recall test) × 2 (criterion level) mixed factor ANOVA, the interaction approached significance, \( F(1,43) = 3.97, \text{MSE} = 1.76, p = .053, \eta_p^2 = .08 \). Although we are primarily interested in determining if the effects of criterion level on associative memory are asymmetric, a difference at any criterion level would be sufficient for showing asymmetry. For instance, for Criterion 5 items, performance was greater on the forward cued recall test \( (M = 35.5\%) \) versus the backward cued recall test \( (M = 22.4\%) \) \[ t(43) = 1.96, p = .029, d = .60 \]. This provides additional evidence that high levels of retrieval success may produce asymmetric gains in associative memory.

Did the same pattern obtain for target memory versus cue memory? To evaluate target and cue memory, we included both free recall and recognition measures. Outcomes on the free recall tests are plotted in Fig. 1. A 2 (type of free recall test) × 5 (criterion level) mixed factor ANOVA revealed a main effect of criterion level, \( F(4,184) = 6.38, p = .001, \eta_p^2 = .11 \).

\[ t(43) = 1.96, p = .029, d = .60 \]
MSE = .85, p < .001, $\eta_p^2 = .12$. The overall performance difference between target memory versus cue memory was not significant, $F(1,46) = 1.96$, MSE = 3.54, $p = .168$, $\eta_p^2 = .04$. Paralleling the pattern in cued recall, we observed greater qualitative gains in target memory versus cue memory (11% versus 6%); however, the interaction between criterion level and recall test format was not significant ($F < 1$). When including only the two extreme criterion levels (criterion level 1 and criterion level 5) in a 2 (type of cued recall test) $\times$ 2 (criterion level) mixed factor ANOVA, the interaction approached significance, $F(1,46) = 3.54$, MSE = .75, $p = .066$, $\eta_p^2 = .07$; for Criterion 5 items, performance was greater for target memory ($M = 16.9\%$) versus cue memory ($M = 10.0\%$) ($t(46) = 1.71$, $p = .047$, $d = .51$).

Next, consider performance on the recognition tests of cue and target memory. For each individual, we calculated corrected recognition as percentage of hits minus percentage of false alarms (hits and false alarms are reported in Table 1). Mean corrected recognition for each test is reported in Table 1. To maintain counterbalanced assignment of participants in each recall group to the recognition tests, some participants necessarily were exposed to information on the recall test that would facilitate performance on the recognition test. Accordingly, we excluded data from seven individuals who completed forward cued recall followed by cue recognition and from eight participants who completed backward cued recall followed by target recognition. A 2 (type of recognition test) $\times$ 5 (criterion level) mixed factor ANOVA revealed a main effect of criterion level, $F(4,180) = 8.02$, MSE = .01, $p < .001$, $\eta_p^2 = .15$, and a main effect of recognition test format, $F(1,45) = 9.47$, MSE = .22, $p = .004$, $\eta_p^2 = .17$. However, unlike the pattern observed in the free recall data, the
gains observed across criterion levels actually favored cue memory versus target memory (15% versus 9%), although the interaction between type of recognition test and criterion level was not significant, \((F < 1)\). Including only the two extreme criterion levels (criterion level 1 and criterion level 5) in a 2 (type of cued recall test) \(\times\) 2 (criterion level) mixed factor ANOVA, the interaction was not significant, \((F < 1)\). Thus, the free recall and recognition measures of cue and target memory yielded somewhat different patterns. One possibility is that these two measures are differentially susceptible to the influence of associative memory (considered further in Experiment 2).

In sum, a marginal difference in gains favored forward versus backward recall in cued recall and target versus cue memory in free recall. These results suggest that the effects of criterion learning on associative memory are symmetric. However, the pattern of gains on the recognition measures provides some preliminary evidence against associative symmetry, as there were not greater gains in target recognition versus cue recognition (with a trend in the opposite direction). Experiment 2 further explored whether the effects of criterion level on associative memory are symmetric or asymmetric.

**Experiment 2**

Experiment 2 was designed to overcome three potential concerns with the free recall tests used in Experiment 1. First, performance was approaching the floor, limiting interpretability. Second, output monitoring may have caused participants to incorrectly withhold responses due to uncertainty regarding whether the recalled word represented a cue or target. To improve free recall performance and eliminate any potential confusion, we combined the free recall of cues and free recall of targets into a single free recall test for all words (i.e., participants were instructed to recall all the words, regardless of whether they were cues or targets).

Third, and most important, the patterns observed in free recall may have been influenced by associative memory. When searching memory for targets, an individual may intentionally or unintentionally retrieve some of the cue words and then use the associative links to access the associated targets. Similarly, individuals attempting to retrieve cue words may recall target words and use the associative link to access the cues. However, if the associative link is weaker in the backward direction, participants would be less likely to access cue words from retrieved targets than targets from retrieved cues. Given that the free recall tests for cues and targets were administered between-participants in Experiment 1, we cannot evaluate the potential influence of associative memory on the free recall measures. Administering the free recall test within-participants in Experiment 2 permits secondary analyses to explore this possibility (described further below).

**Method**

Eighty-seven undergraduates participated for course credit. The materials, design, and procedure were identical to Experiment 1, except for one difference pertaining to the free recall groups. In Experiment 2, we administered only one free recall test, in which participants were instructed to recall as many English words as possible (i.e., both cues and targets). We excluded data for five participants who did not reach at least 75% criterion for all items within the allotted time during Session 1.

**Results**

To revisit, evidence that the effects of criterion level on associative memory are symmetric would involve either (1) equivalent gains in forward and backward cued recall and equivalent gains in target memory and cue memory, or (2) greater gains on the forward versus backward cued recall test and greater gains for target memory versus for cue memory. First, consider forward versus backward cued recall performance (see Fig. 1). A 2 (type of cued recall test) \(\times\) 5 (criterion level) mixed factor ANOVA revealed a main effect of criterion level \([F(4,200) = 22.82, \text{MSE} = 1.87, p < .001, \eta^2 = .313]\); the main effect of test type was not significant, \([F(1,50) = 1.87, \text{MSE} = 24.29, p = .178, \eta^2 = .036]\). Of greatest interest, the interaction was significant, \([F(4,200) = 2.51, \text{MSE} = 1.87, p = .043, \eta^2 = .048]\). Thus, the same qualitative pattern as in Experiment 1 emerged, with greater gains in forward versus backward cued recall (31% versus 16%). These results would suggest that the effects of criterion level on associative memory are asymmetric, unless the same pattern is observed for target memory versus cue memory.

Free recall of cues and free recall of targets is reported in Fig. 1. A 2 (type of free recall test) \(\times\) 5 (criterion level) mixed factor ANOVA revealed a main effect of criterion...
level, $F(4,116) = 9.89, \text{MSE} = 1.48, p < .001, \eta^2_p = .25$. The overall performance difference between target recall and cue recall was significant, $F(1,29) = 4.95, \text{MSE} = 972, p = .034, \eta^2_p = .15$. However, in contrast to the pattern in cued recall, the interaction was not significant ($F < 1$). The gains across criterion level for target memory and cue memory were similar (14% versus 11%).

One possibility is that the recall of target words interfered with their recall of cue words. If participants had greater access to target information in memory, those items may have been recalled first in their recall output. If so, then the cue words would receive greater output interference and thus be more difficult to recall as compared to the targets. To examine this, we divided each participant’s recall output into quartiles. Then, we calculated the proportion of target words recalled within each quartile. Results of a one-way ANOVA suggested that the proportion of targets recalled was approximately the same across each quartile (55%, 58%, 57%, and 62%, respectively; $F < 1$), suggesting minimal output interference.

As foreshadowed above, we also conducted secondary analyses of free recall to explore the extent to which associative memory may have influenced free recall of targets and cues. First, we identified cases in which participants recalled both the cue word and target word of a given word pair (e.g., recalling both ‘queen’ and ‘journal’ for the word pair ‘queen–journal’). Next, we determined the number of cases in which the cue and target were recalled adjacent (e.g., ‘queen’ followed immediately by ‘journal’) versus non-adjacently (e.g., other items were recalled between ‘queen’ and ‘journal’). A significantly greater number were recalled adjacent versus non-adjacently, (2.53 versus 0.88). $t(29) = 3.76, p = .001$, suggesting that associative memory contributed to free recall (cf. associative chaining effects on free recall of word lists; Kahana, 2002). Importantly, was the influence of associative memory on free recall symmetric or asymmetric? Of the cue words recalled, 36.2% were recalled immediately by their associated target, whereas only 22.3% of recalled targets were followed immediately by their associated cue, $t(25) = 1.80, p = .042$, suggesting that associative memory had an asymmetric influence on free recall of targets versus free recall of cues. In contrast, an associative symmetry account would reasonably have predicted an equivalent likelihood of traversing from cue to target as from target to cue, given the assumption of one bidirectional link.

Outcomes on the recognition tests (reported in Fig. 1) provide further evidence that the effects of criterion level on associative memory are asymmetric. As in Experiment 1, we calculated corrected recognition as percentage of hits minus percentage of false alarms. (For reasons noted in Experiment 1, we excluded data from ten individuals who completed forward cued recall followed by cue recognition and from eight participants who completed backward cued recall followed by target recognition). A 2 (type of recognition test) × 5 (criterion level) mixed factor ANOVA revealed a main effect of criterion level [$F(4,144) = 15.12, \text{MSE} = .01, p < .001, \eta^2_p = .11$] and a main effect of recognition test format [$F(1,36) = 6.07, \text{MSE} = .16, p = .019, \eta^2_p = .14$]. Importantly, the gains with criterion level were not greater for target memory versus cue memory (the interaction was not significant, $F < 1.47$). Indeed, the gains observed across criterion level numerically favored cue memory versus target memory (25% versus 14%).

In sum, Experiment 2 provided stronger evidence that the effects of criterion level on associative memory are asymmetric. First, we observed greater gains in forward versus backward cued recall, but similar gains in target memory versus cue memory as measured via free recall. Furthermore, secondary analyses showed that free recall was influenced by associative memory (more adjacent versus non-adjacent pairs were recalled), and that this influence was asymmetric (more word pairs were recalled from cue-to-target versus target-to-cue). Second, the recognition measures showed numerically greater gains for cue memory versus target memory. Thus, the greater gains in forward versus backward cued recall cannot easily be attributed to greater gains in target memory versus cue memory, implying asymmetric gains in associative memory.

**Experiment 3**

The conclusion that the effects of criterion level on associative memory are asymmetric rests in part on the consistent finding that the recognition measures did not show greater gains in target versus cue memory, which is problematic for an associative symmetry interpretation given consistently greater qualitative gains in forward versus backward cued recall. However, a potential methodological concern is that the recall tests always preceded the recognition tests. To ensure that the pattern of recognition was not qualitatively influenced by the preceding recall test, we only administered recognition tests in Experiment 3.

**Method**

Seventy-two undergraduates participated for course credit. Materials, design, and procedure were identical to Experiment 1 and 2, except that there were no recall tests. Instead, participants only completed a recognition test, either cue recognition or target recognition (or associative recognition; see Footnote 1). We excluded data for four participants who did not reach at least 75% criterion for all items during Session 1.

**Results**

Corrected recognition is reported in Fig. 2 (hits and false alarms are reported in Table 1). A 2 (type of recognition test) × 5 (criterion level) mixed factor ANOVA on corrected recognition revealed a main effect of criterion level, $F(4,168) = 20.23, \text{MSE} = .02, p < .001, \eta^2_p = .33$, as well as a main effect of recognition test format, $F(1,42) = 9.12, \text{MSE} = .16, p = .004, \eta^2_p = .18$. As in the previous experiments, the interaction between type of recognition test and criterion level was not significant ($F < 1$), and the numerical trends favored cue memory versus target memory (25% versus 18%). These results replicate Experiments 1 and 2 and provide additional evidence suggesting that the
effects of criterion level on associative memory are asymmetric, as the gains were not greater for target memory versus cue memory.

Experiment 4

The conclusion that the effects of criterion level on associative memory are asymmetric rests in part upon the consistent finding that the effects of criterion level did not differ for cue and target recognition (and were numerically greater for cues). However, a potential concern is that the absence of an interaction between criterion level and recognition performance may reflect scale dependency (e.g., Loftus, 1978). In brief, recognition performance presumably reflects the state of an underlying memory representation. Although test performance and the underlying memory representation are assumed to be monotonically related, the specific function relating these two is not known (e.g., linear, curvilinear, sigmoidal). If the function is not linear, then more caution is needed when interpreting interactions or the lack thereof when comparing test performance in two conditions that are at different parts of the scale. For example, 15% improvements in the lower versus upper parts of the performance scale (e.g., from 5% to 20% versus from 80% to 95%) may not necessarily reflect the same degree of improvement in the underlying memory representations. In contrast, comparisons that involve overlap of performance at some point on the scale support more confident conclusions about the underlying memory representations.

Accordingly, the purpose of Experiment 4 was to minimize concerns about scale-dependency by including conditions in which cue and target recognition were at similar levels on the performance scale. Specifically, to boost cue memory, we implemented a pre-exposure condition for half of the cue words, in which cues were presented for initial study prior to criterion learning of the word pairs. In both the pre-exposure and no-pre-exposure conditions, we expected better forward versus backward cued recall performance (as observed in Experiments 1 and 2). Importantly, Experiment 4 also established that there are no differences between gains in target memory versus cue memory, even when cue recognition and target recognition performance are at similar parts of the scale.

Method

One hundred fifteen undergraduates participated for course credit. Materials, design, and procedure were identical to Experiment 1, with the following exceptions. First, due to the increase in the number of possible conditions with the addition of the pre-exposure manipulation, we eliminated two intermediate criterion levels from the initial learning session to simplify the design and to maintain a reasonable number of items in each remaining condition. Thus, participants correctly recalled items one, three, or five times. Second, to equate the number of items assigned to each criterion level and pre-exposure condition, we reduced the number of items from 50 word pairs to 48 word pairs (i.e., 6 lists of 8 word pairs). Assignment of list to pre-exposure condition and criterion level was approximately counterbalanced across participants. Third, we administered a pre-exposure phase at the beginning of the experiment in which the cue words from half of the word pairs (24 cue words) were presented for an initial study trial one at a time for 4 s each. After the first block of pre-exposure trials, the procedure was repeated for two additional blocks of pre-exposure trials. Presentation order was randomized during each block of pre-exposure trials. Finally, given evidence from Experiment 2 that free recall of cues and targets was influenced by associative memory, we eliminated the free recall tests. We excluded data for one participant who did not reach at least 75% criterion for all items during Session 1.

Results

Cued recall outcomes are reported in Fig. 3. A 2 (pre-exposure or no-pre-exposure) × 2 (type of cued recall test) × 3 (criterion level) mixed factor ANOVA revealed a main effect of criterion level \(F(2,114) = 85.16, MSE = 1.32, p < .001, \eta^2_p = .60\) and a significant 2-way interaction between type of cued recall test (forward versus backward) and criterion level \(F(2,114) = 8.89, MSE = 1.32, p < .001, \eta^2_p = .14\). Thus, the same qualitative pattern as in Experiments 1 and 2 emerged, with greater gains in forward versus backward cued recall (32% versus 16%). There was no main effect for pre-exposure condition \(F(1,57) = 1.40, MSE = 1.31, p = .242, \eta^2_p = .02\) nor any interactions involving pre-exposure condition \(Fs < 1.41\). These results replicate Experiments 1 and 2 by showing greater gains in forward versus backward cued recall.

Corrected recognition outcomes are reported in Fig. 4 (hits and false alarms are reported in Table 2). A 2 (pre-exposure or no-pre-exposure) × 2 (type of recognition test) × 3 (criterion level) mixed factor ANOVA revealed a main effect of criterion level \(F(2,106) = 39.41, MSE = .02, p < .001, \eta^2_p = .43\) and pre-exposure condition \(F(1,53) = 4.57, MSE = .03, p = .037, \eta^2_p = .08\). Additionally, the interaction between pre-exposure condition and type of recognition test was significant \(F(1,53) = 5.64, MSE = .03, p = .021, \eta^2_p = .10\), as well as the interaction between pre-exposure condition and criterion level \(F(2,106) = 5.37, MSE = .02, p = .006, \eta^2_p = .09\). There was also a significant 3-way interaction \(F(2,106) = 4.76, MSE = .02, p = .010, \eta^2_p = .08\) (all other \(Fs < 3.62\). To
explore the 3-way interaction, we conducted separate 2 (type of recognition test) × 3 (criterion level) ANOVAs for the pre-exposure and no pre-exposure conditions. The interaction between type of recognition test (cue versus target) and criterion level was significant in the no-pre-exposure condition \[ F(2,106) = 5.10, \text{MSE} = .02, p = .008, \eta^2_p = .09 \], but not in the pre-exposure condition \( F < 1 \). These results confirm greater gains for cue memory versus target memory on the recognition measures in the no-pre-exposure condition (26% versus 11%) and no significant difference in gains in the pre-exposure condition (11% versus 15%), neither of which is consistent with a symmetry account.

In sum, Experiment 4 showed once again that gains were greater in forward versus backward cued recall. Importantly, gains favored forward versus backward cued recall regardless of whether cue memory and target memory gains were equivalent (in the pre-exposure condition) or whether gains were greater for cue versus target memory (in the no pre-exposure condition). Therefore, the difference in forward versus backward cued recall performance cannot be attributed to differences in cue memory or target memory, providing additional evidence that the effects of criterion level on associative memory are asymmetric.

Table 2

Mean percentage of hits and false alarms as a function of criterion level and type of recognition test in Experiment 4.

<table>
<thead>
<tr>
<th></th>
<th>Hits for criterion level N</th>
<th>False alarms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Cue recognition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pre-exposure</td>
<td>56.9 (4.8)</td>
<td>78.2 (3.5)</td>
</tr>
<tr>
<td>Pre-exposure</td>
<td>77.8 (3.6)</td>
<td>83.3 (3.2)</td>
</tr>
<tr>
<td><strong>Target recognition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pre-exposure</td>
<td>78.1 (3.1)</td>
<td>87.1 (3.2)</td>
</tr>
<tr>
<td>Pre-exposure</td>
<td>77.7 (3.9)</td>
<td>87.9 (2.3)</td>
</tr>
</tbody>
</table>

*Note: False alarm rates in each condition are identical because only one set of lures was used.*

Fig. 3. Mean final cued-recall performance as a function of criterion level, type of cued recall test, and pre-exposure condition in Experiment 4. Performance is reported in percentages, and error bars report standard error of the mean.

Fig. 4. Mean final recognition performance as a function of criterion level, type of recognition test, and pre-exposure condition in Experiment 4. Performance is reported in percentages, and error bars report standard error of the mean.

General discussion

A wealth of research has established that successful retrieval practice enhances subsequent memory, with recent research establishing monotonic increases in memory as criterion level increases (e.g., Vaughn & Rawson, 2011). The current work explored the nature of criterion level effects on subsequent memory, with particular focus on the gains in associative memory across criterion levels. Across experiments, the overall pattern of results weighed against symmetric associative gains under conditions of criterion learning and instead indicate that the effects of criterion level on associative memory are asymmetric.

Although these outcomes indicate associative asymmetry and are consistent with the original findings of Vaughn and Rawson (2011), these results diverge from the common finding of associative symmetry within the associative memory literature (see Kahana, 2002). We discuss two possible factors which may have contributed to these diverging results. First, we utilized a one-week retention interval, whereas most prior studies investigating associative symmetry used shorter delays between initial learning and final tests. One plausible explanation is that differences between forward and backward associative links may not be readily apparent on short retention intervals but may emerge after a delay. This possibility is consistent with assumptions of the bifurcated distribution model (BDM; Kornell, Bjork, & Garcia, 2011). According to BDM, items gain memory strength from testing (i.e., successful retrieval) as well as from re-exposure (i.e., restudying the items). Items gain the most memory strength when they are tested and correctly recalled, whereas items gain no memory strength when they are tested and not correctly recalled. As such, tested items become bifurcated in memory (with items correctly recalled having high memory strength and items not correctly recalled having low memory strength). Items that are only restudied gain a moderate amount of memory strength (i.e., more strength than would occur after a failed test but less strength than would occur after a correct test) and are represented by a normal distribution of memory strength (i.e., a non-bifurcated distribution). Items are assumed to be correctly recalled on a final retention test if they have sufficient memory strength to cross the recall threshold. BDM was originally proposed to explain why performance differences between testing versus restudy often do not emerge after a short retention interval or even favor restudied items (e.g., Hogan & Kintsch, 1971; Thompson, Wenger, & Bartling, 1978). According to BDM, after short retention intervals, the intermediate levels of memory strength for many restudied items is sufficient to cross the recall threshold. However, all items lose memory strength at an equal rate over time (i.e., an equal rate of forgetting occurs for both restudied and tested items). As such, at longer retention intervals, restudied items may no longer have sufficient memory strength to cross threshold, in contrast to tested items that accrued high levels of memory strength from successful retrieval practice (e.g., Wheeler, Ewers, & Buonanno, 2003). By extension to the present research, one possibility is that forward associations accrue large increments in memory strength from successful retrieval practice, whereas backward associations accrue only modest increments in memory strength during restudy. If so, both forward and backward associations may be sufficiently strong to cross threshold after a short delay, whereas the stronger forward associations would be favored at a longer retention interval.

Second, prior research on associative symmetry has typically involved a fixed number of practice trials rather than criterion learning, which is the condition of interest here. With a fixed number of trials, some items will not be correctly recalled during practice (likely the most difficult items) and many items will only be correctly recalled once. By comparison, forward and backward cued recall for Criterion 1 items in the current experiments tended to be similar, suggesting that evidence for associative asymmetry may only emerge with higher levels of retrieval success. As reviewed earlier, the bulk of prior research on associative memory suggests associative symmetry. Of the few studies that did suggest associative asymmetry, it is interesting to note that these studies involved overlearning (e.g., Giurintano, 1972; Vaugh, 1970; Wollen, Fox, & Lowry, 1970). Interpretation of these outcomes require some caution because they did not equate or include measures of target and cue accessibility, but their findings are consistent with the current outcomes. One plausible explanation why associative asymmetry may only emerge with higher levels of retrieval success pertains to item difficulty. With a relatively low criterion level (e.g., Criterion 1), items correctly recalled on the final cued recall test may consist primarily of easier items. Difficult items are presumably recalled more frequently at higher versus lower criterion levels (contributing to the increases in overall recall observed at higher criterion levels). One possibility is that differences in forward and backward associative strength are relatively minimal for easy items, but these differences become more apparent for difficult items. If so, as the proportion of difficult items contributing to recall increases across criterion levels, a difference between forward and backward cued recall would also be expected to emerge at higher criterion levels. Of course, this account is speculative, but it does suggest interesting directions for future research.

In addition to addressing the theoretical question concerning the nature of criterion level effects on associative memory, the current outcomes also have important implications for testing effect theories more generally, in that they impose potential constraints on theories explaining the benefits of retrieval practice. As mentioned above, the mediator effectiveness hypothesis and the elaborative retrieval hypothesis make no explicit claims about associative symmetry, but extensions reasonably follow from both accounts. As outlined earlier, the mediator effectiveness hypothesis attributes the benefits of retrieval practice to generating more effective mediators. According to the elaborative retrieval hypothesis, retrieval practice activates related semantic information which can later serve as additional retrieval cues to access targets. Both theories appear to be able to account for associative asymmetry.
results. For instance, mediators tend to be asymmetric, often linking a cue word to a particular target word via phonetic similarity (e.g., wingu – wing – cloud). Testing memory in a backward direction (e.g., cloud – ?) would presumably not activate the previously generated mediator due to lack of phonetic similarity (e.g., wing would likely not be activated when cued with cloud). Likewise, semantic information activated during testing (e.g., morning breakfast, shower, commute, tired light) would likely only be reactivated during a forward test of memory (e.g., morning – ?) and not necessarily during a backward test of memory (e.g., light – ?). Again, although these two accounts are silent with respect to associative symmetry, both accounts appear to be able to readily account for findings of associative asymmetry (e.g., a consistent advantage for forward versus backward cued recall).

In contrast, it is less obvious how other retrieval practice theories can account for the patterns of associative asymmetry demonstrated here. In particular, descriptive strength-based models that do not posit underlying explanatory mechanisms are less well equipped to account for the current outcomes. For instance, according to the retrieval effort hypothesis (Pyc & Rawson, 2009), retrieval success boosts subsequent memory to a greater extent when successful retrieval involves more versus less effort. However, this hypothesis is a descriptive account that does not explain what effort reflects nor why it improves memory, and thus it is not well equipped to explain why criterion level has asymmetric effects on associative memory. Thus, this account is silent concerning differential gains in forward versus backward associative memory without further specification of the underlying mechanisms that give rise to the patterns described by these accounts.

In conclusion, the overall pattern of results across four experiments converges on the theoretical conclusion that the effects of criterion level on associative memory are asymmetric. The current work represents an important foundation for future research to further explore the nature of criterion level effects in particular and the memorial consequences of criterion learning more generally.

References


