Who benefits from memory training?

David Bissig
Wayne State University

Cindy Lustig
University of Michigan

In press, *Psychological Science*

Address correspondence to:
Cindy Lustig
Department of Psychology
University of Michigan
530 Church Street
Ann Arbor, MI 48109-1043

Ph: 734 647 6925

Email: clustig@umich.edu
Abstract

Cognitive training programs can have significant benefits. However, their efficacy is often reduced for those of advanced age or lower cognitive ability. Using older adult participants, we examined the role of self-initiation of cognitive control in a training program that targets recollection memory. Relative time spent on an open-ended, intentional encoding task that requires the self-initiation of cognitive control was highly predictive of improvement in the training task, and fully accounted for individual differences related to age and crystallized intelligence. Analyzing training programs from the perspective of cognitive theory may help understand how these programs are having their effects and suggest ways to optimize such programs for those who need them most.
Who benefits from memory training?

Training programs are an increasingly popular way of treating the cognitive deficits associated with a wide variety of conditions, including schizophrenia, head trauma, multiple sclerosis, normal aging, and Alzheimer’s disease (Ball et al., 2002; Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Clare & Woods, 2004; Jennings & Jacoby, 2003; Jennings, Webster, Kleykamp, & Dagenbach, 2005; Loewenstein, Acevedo, Czaja, & Duara, 2004; McGurk, Mueser, & Pascaris, 2005; Prosiegel & Michael, 1993; Rund & Borg, 1999; Twamley, Jeste, & Bellack, 2003; Verhaeghen, Marcoen, & Goossens, 1992; Whyte, 2006). The benefits of training can extend beyond the trained task and lead to significant changes in both behavior and brain function (Ball et al., 2002; Jennings et al., 2005; Loewenstein et al., 2004; Lustig & Buckner, 2004; Nyberg et al., 2003). However, it is often unclear what factors contribute to the success of a training program. What processes are being trained? Why do some individuals benefit more than others?

This paper asks if individual differences in controlled processing, an important concept in basic theories of cognition, can account for individual differences in memory training. Training programs are successful at the group level, but individual differences in the degree of improvement are quite large. Unfortunately, the benefits of training are often smallest for those who need them most: Both lower initial cognitive status and advanced age are associated with smaller improvements (Verhaeghen et al., 1992; Yesavage, Sheikh, Friedman, & Tanke, 1990). Understanding the reasons for these individual differences could guide the design of future training programs to better benefit disadvantaged groups.

Individual differences in controlled processing contribute to individual differences on many cognitive tasks. The production deficit hypothesis suggests that many of older adults’ cognitive deficits stem from failures to self-initiate controlled, effortful processes that support successful performance (Craik & Byrd, 1982). By this view, deficits may be remediated by appropriate environmental support or instruction. For example, older adults fail to activate brain regions associated with successful memory if given open-ended, intentional learning instructions (“Memorize the words”), but activate these regions as much as do young adults if instructed to perform a semantically-based deep encoding task that supports later memory (Logan, Sanders, Snyder, Morris, & Buckner, 2002).

Recent models extend the production deficit hypothesis to differences in when and how control is engaged. By the dual mechanisms of cognitive control account, variation in cognitive tasks is linked to whether control is engaged in a proactive or reactive fashion (Braver, Gray, & Burgess, in press). High-ability individuals (e.g., young adults with high fluid intelligence) emphasize proactive control to maintain task goals that “set the stage” for successful cognition. In contrast, low-ability young adults and older adults engage control in a more reactive fashion, in response to immediate task demands. Likewise, aging is associated with a “load shift” from early processes that constrain retrieval to relevant events, to later processes that evaluate information after retrieval (Velanova, Lustig, Jacoby, & Buckner, in press). Neuroimaging suggests an even more extreme version of the load-shift model in some cases, with older adults failing to activate control-related frontal regions under intentional encoding instructions.
but activating them more than young adults at retrieval, possibly in an attempt at
compensation (Head, Lustig, Isom, & Buckner, 2006).

The training program used here focuses on recollection, or deliberate,
consciously-controlled memory retrieval processes considered to be distinct from more
automatic, familiarity-based processes (Jacoby, 1991). Recollection requires more
effort and cognitive control, and involves specific, analytic details such as source (“Was
that from the news or an advertisement?”) or modality (“Did I hear it or read it?”). Recollection’s control-demanding,
effortful aspects lead to its frequent failure in those with reduced cognitive control, such as older adults, brain-damaged patients, or
individuals with low working memory spans (Dockree et al., 2006; Jennings & Jacoby,
1993; Oberauer, 2005). These failures can have social and practical consequences. A
classic example is the older individual who consistently re-tells a favorite story to a
group of acquaintances, failing to recollect having done so on previous occasions. There can be more serious consequences, as when scam artists introduce false
information (e.g., about pricing) into a conversation to increase its familiarity, and
individuals fail to recollect the lower price agreed upon earlier (Jacoby, 1999).

Recollection failures can be reduced by a training program that gradually
increases the demand to respond on the basis of controlled, analytic processes
(Jennings et al., 2006; Jennings & Jacoby, 2003; Jennings et al., 2005). Participants
complete multiple sessions during which they learn lists of words and then must
discriminate studied words from unstudied lures. Lures are repeated during the test.
Recollection is required to correctly identify repeated items as lures, and avoid the urge
to designate them as “studied” based on familiarity. Recollection demands are
gradually increased by increasing the number of items between repetitions.

The benefits of this training program have shown transfer to other tasks
(Jennings et al., 2005), but only those that also require discrimination between correct
versus incorrect but familiar items (e.g., the n-back task, which requires participants to
identify whether a probe stimulus occurred exactly “n” trials previously in a series). This
pattern supports the idea that training benefits specific processes, rather than reflecting
a generic improvement due to increased attention or social stimulation (c.f., Ball et al.,
2002). Both healthy and cognitively-impaired individuals showed training-related
improvements, but with large individual differences (Jennings et al., 2003, 2005, 2006).
Some individuals reached maximal success, ultimately performing perfectly even with
40 items between lure repetitions. Others showed almost no improvement despite 28
sessions of training.

What drives these individual differences in training efficacy? The most intuitive
answer is that they reflect processes engaged at retrieval, especially those involved in
discriminating studied from repeated, unstudied items. However, processes engaged at
encoding can have important downstream effects, especially for recollection. Controlled
processing at encoding (especially deep, meaning-based processing) typically provides
better support for later retrieval (Craik & Lockhart, 1972). Reaction time, accuracy, and
brain activation at retrieval are strongly influenced by prior encoding processes
(Velanova et al., 2003). Of particular interest, encoding processes influence the
retrieval-test processing of lure items, not just studied items, although this relationship
may break down with age (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu,
Velanova, & Rhodes, 2005). Both encoding and retrieval involve cognitive control, but
encoding may require more self-initiation of control and thus be more vulnerable to disruptions (e.g., from divided attention). By contrast, retrieval appears to be more “obligatory” and takes priority when competing with other tasks (Anderson, Craik, & Naveh-Benjamin, 1998; Naveh-Benjamin, Craik, Guez, & Dori, 1998).

We asked whether differences in controlled processing could account for individual differences in recollection training efficacy. This training program uses intentional encoding, an open-ended task that requires the self-initiation of controlled encoding strategies to support later memory. By contrast, although the retrieval task also requires control – especially to discriminate between studied words and familiar, but unstudied, repeated lures – its structure (a recognition task) makes those demands more obvious, reducing the demand for self-initiation. We hypothesized that individual differences in training efficacy related to age and initial ability would be related to the self-initiation of cognitive control, as measured by the tendency to emphasize processing at encoding rather than retrieval.

Method

Participants. Participants were 19 adults (mean age 74.5 yrs, SD = 6.1; mean years of education 18.1, SD = 3.2). None had medications or conditions affecting attention, memory, or movement. An additional participant’s data were dropped because of low accuracy on studied words (59%; a 2.5 SD outlier from the group mean).

Materials and procedure. The recollection training procedure was modified (from Jennings & Jacoby, 2003; Jennings et al., 2005) to allow for self-paced presentation at all stages and feedback after every test trial. Stimuli were chosen from the English lexicon project (Balota et al., 2002; mean length 5.76 letters, range 3-9 letters; mean frequency 20,487 out of 131 million; length and frequency balanced across lists and conditions) and presented using E-Prime. Response times were collected via keypress.

Participants completed four study-test sessions per day for seven days, scheduled over the course of two weeks. During the study phase, participants studied 30 individually-presented words. During the test phase, they performed an old-new recognition test, requiring discrimination of studied words from unstudied lures. Unstudied lures were repeated within the test list. Each recognition test had 90 items, pseudorandomly intermixed: studied words, first presentations of unstudied lures, and second presentations of unstudied lures. Each response on the retrieval test was followed by a feedback screen indicating accuracy (correct or incorrect) and trial type (studied, new, or repeated).

The critical manipulation was the lag between lure repetitions. Participants started at an easy level, with half the lures repeated after one intervening word, and half the lures repeated after two intervening words. (See Figure 1). Difficulty gradually increased, with possible lag intervals of: 1&2, 1&3, 2&4, 2&8, 4&12, 4&16, 8&20, 8&24, 12&28, 12&32, 16&36, and 16&40. Participants were thus always working at one relatively easy interval and one that might be more challenging. Lag level was increased when the participant reached criterion at the current level (96% correct on the more challenging interval for the first four levels (up to 2&8); 93% correct for higher levels).
Training Memory

Participants completed informed consent prior to the first training session and also completed a health and demographics questionnaire, the Extended Range Vocabulary Test (ERVT, Version 3, Educational Testing Services, 1976), the Mini Mental State Evaluation (MMSE, Folstein, Folstein, & McHugh, 1975), and other forms.

Results

Ranking. Participants were ranked (1 = best, 19 = worst) by final lag level (e.g. better rank for a participant who achieved lag level 4&16 by Session 28 versus one who had only achieved level 2&8). Eight participants achieved the maximum lag level, 16&40. “Ties” between participants at the same level at the end of the experiment were resolved by assigning the better rank to the participant who had reached criterion earlier in training (e.g., reaching level 16&40 in Session 12 vs Session 15). Remaining ties were broken by assigning the better rank to the participant with better overall correct rejection of repeated lures.

Because two-thirds of the items are unstudied, a potential concern is that participants could reach a high ranking by simply responding “unstudied” to most items, rather than using memory. This was not the case: Rank correlated positively with accuracy on studied items, r = .85, p_{rep} > .999.

Replication of age and ability effects. In separate analyses, age was negatively related to performance, accounting for 25% of the variance in rank (b* = -.54, p_{rep} = .95). Crystallized intelligence (ERVT score), was positively related to rank, accounting for 22% of its variance, b* = .51, p_{rep} = .94. See Figure 2, panels A & B. The MMSE dementia scale showed similar trends (adjusted R^2 = .17, b* = .41, p_{rep} = .89) but did not reach significance due to ceiling effects (range 25-30; all but 3 participants scored 28 or

Figure 1. Illustration of study and test conditions. Participants first engaged in self-paced intentional encoding of individually-presented words (left). At retrieval, they indicated whether each word was from the study list or was an unstudied lure. Unstudied lures (e.g., SILVER, RIFLE) were repeated within the test list. Repetitions within a list occurred at either a short or a long lag. The list illustrated here had a short lag of 1 intervening item (between the first and second presentation of RIFLE) and a long lag of 2 intervening items (between the first and second presentation of SILVER). Difficulty was increased by increasing lag between items once participants reached criterion performance (see text).
above). These patterns replicate previous findings that age and lower initial cognitive ability are negatively related to training efficacy (Verhaeghen et al., 1992; Yesavage et al., 1990).

![Figure 2](image.png)

**Figure 2.** Age and crystallized intelligence as predictors of rank before and after accounting for proportional study time. Left panels (A & B) are the scatterplots for rank on the training task (1 = best, 19 = worst) as a function of age (panel A; range 67-93 yrs) or vocabulary score (panel B; range 16-46), a measure of crystallized intelligence. Right panels (C & D) show the reduction of these relations after accounting for proportional study time.

*Time spent on encoding versus retrieval.* No participant spent more time on retrieval trials (M = 1.5 s, SD = .38) than on study trials (M = 5.7s, SD = 4.07). Study and retrieval time were not related to one another (r = .07) and showed opposite-sign relations to rank (r = -.76 for study; r = .20 for retrieval). To measure the relative amount of time spent on study versus retrieval, we calculated a proportional index: (mean study time – mean retrieval time) / (mean study time + mean retrieval time). Proportional study time (PST) was preferred over raw study time because the latter
training Memory might reflect a general tendency to be slow, rather than the emphasis on encoding. PST was positively related to performance (Figure 3), accounting for 67% of the variance in rank ($b^* = .83, p_{rep} > .999$). It showed a small but significant increase over sessions, consistent with the idea that training enhanced an emphasis on encoding (mean slope = .004, SD = .007, $t(18) = 2.61, p_{rep} = .95$).

Does PST account for age and ability effects? PST was negatively related to age ($r = - .62, p_{rep} = .999$) and positively related to ERVT ($r = .53, p_{rep} = .95$). Could it account for age and ERVT relationships with rank? A combined model with PST entered first and age second accounted for 69% of the variance in rank, and age was not a significant contributor ($R^2$ change = .001, $b^* = -.04, p_{rep} = .55$). Similar patterns were found for ERVT ($R^2$ change = .001, $b^* = .10, p_{rep} = .66$), and MMSE ($R^2$ change = .02, $b^* = .17, p_{rep} = .79$). (Figure 2, Panels C & D.) The standardized regression coefficient ($b^*$) for PST changed very little depending upon what other variables were in the model (from .83 to .81 if age was included, .77 if ERVT or MMSE was included).  

Figure 3. Proportional study time strongly predicts performance. Participants who spent relatively more time per item on encoding than on retrieval performed better on the training task. See text for details.
Contributions at retrieval? Analyses so far support the hypothesis that emphasizing encoding processes improves training efficacy. An exploratory analysis for retrieval-based effects found proportional retrieval time for repeated (familiar) versus new items only marginally predicted rank (adjusted $R^2 = .11$, $b^* = .40$, $p_{rep} = .88$). Its slope did not change over sessions, (mean = -.0008, SD = .006, $p_{rep} < .70$).

The retrieval-time relation to rank was primarily driven by long-lag trials: Individuals who spent proportionally more time on long-lag versus short-lag trials had worse ranks (adjusted $R^2 = .21$, $b^* = -.51$, $p_{rep} = .94$). Proportional time spent on long-lag trials did not account for age and ability effects. Age remained a significant predictor ($R^2$ change = .17, $b^* = -.43$, $p_{rep} = .92$); ERVT was marginal ($R^2$ change = .12, $b^* = .37$, $p_{rep} = .87$). Proportional time spent on long-lag items did not correlate with PST, and made a unique contribution to the prediction of rank (omnibus adjusted $R^2 = .73$, $R^2$ change = .07, $b^* = .28$, $p_{rep} = .92$).

Individual differences in retrieval-stage processing might also appear in sensitivity to feedback. Proportional time spent on feedback screens for incorrect versus correct items had a marginally positive influence after accounting for PST ($R^2$ change = .06, $b^* = .33$, $p_{rep} = .91$). This effect should be interpreted cautiously, given the different number of “incorrect” screens for good versus poor performers.

Discussion

What leads to successful memory training? At least for the current procedure, a major influence is the self-initiation of controlled processes that support memory. The relative amount of time individuals spent on an open-ended, intentional encoding task was strongly related to improvement on a recollection training task. Furthermore, relative study time accounted for age and ability differences in training recollection. This result fits well with theories that emphasize group and individual differences in self-initiating cognitive control as a source of performance differences (Braver et al., in press; Craik & Byrd, 1982; Velanova et al., in press). They may also relate to findings of paradoxical increases in age differences after training (e.g., Baltes & Kleigl, 1992). If young adults are more likely to self-initiate and practice successful strategies during training than are older adults, they might well show larger improvements.

Study time served as a measurable predictor for regression analyses, but was presumably a proxy for processes engaged at encoding. Individual differences in encoding strategies correlate with brain activation patterns, which in turn predict memory performance (Kirchhoff & Buckner, 2006). In debriefing, several top performers reported strategies such as creating a narrative using the to-be-remembered words, or relating the words to experiences in their own life. These strategies would lead to the deep, meaning-based encoding that typically supports later memory, and might also create a narrative structure integrating the words from the current study list into a unit (Radvansky, 2005). By contrast, poor performers demonstrated either little insight about strategies (“I just used my memory”) or used superficial strategies such as rote rehearsal.

At first, it may seem counterintuitive that an emphasis on encoding should be the primary influence on performance: The manipulation of difficulty (lag between repeated lures) occurs at retrieval. However, this result is predicted by theoretical perspectives
that consider the degree to which different tasks (intentional encoding vs recognition memory) require the self-initiation of controlled processing, and individual and group differences in how such processes are engaged (Braver et al., in press; Craik & Byrd, 1982; Craik & Lockhart, 1972; Velanova et al., in press). Encoding processes can have important downstream effects for how control is engaged at retrieval (e.g., Jacoby, Shimizu, Daniels et al., 2005). In a false memory procedure, warnings presented at encoding reduced false recognition by older adults, whereas warnings presented at retrieval did not (McCabe & Smith, 2002).

From a practical perspective, the somewhat surprising importance of encoding underscores the value of carefully analyzing training programs to understand the locus of their effects. Many training programs show limited transfer to other tasks despite large improvements on the training task itself, and those benefits that do transfer are typically domain- or process-specific (c.f., Ball et al., 2002). Our data suggest that processes engaged at encoding could be a highly effective target for training recollection.

The link between controlled processing at encoding and training efficacy suggests a potential method of modifying training programs to benefit those who initially do not show improvement. Providing environmental support improves older adults’ performance by reducing the need to self-initiate control (Craik & Byrd, 1982). Older or lower-ability individuals might benefit from environmental support in the form of explicit coaching (e.g., requiring a verbal response that links to-be-remembered words across trials), and gradual shaping to initiate these processes on their own. It remains to be seen whether this modification would be successful, or whether similar modifications would apply to other training programs. However, the general approach of using basic theoretical perspectives to analyze training programs holds promise for understanding how these programs have their effects, and how to improve them.

These data may also be relevant to general questions about recognition memory, particularly when it is important to discriminate between different sources of familiarity (Jacoby, Shimizu, Daniels et al., 2005; Jacoby, Shimizu, Velanova et al., 2005). Jacoby and colleagues have argued that a theoretically important distinction can be drawn between proactively restricting retrieval attempts and memory access to the target source (source-constrained retrieval) and retroactively evaluating information once it has been retrieved (source identification). In their experiments, young adults showed evidence for source-constrained retrieval (better memory for foils on a recognition test following deep, meaning-based encoding as opposed to shallow or intentional encoding), but as a group older adults did not.

In a repeated-lure procedure like the one used here, source-constrained retrieval might be used to limit the familiarity or memory access of lure items. Previous work suggests that better encoding supports better source-constrained retrieval (Jacoby, Shimizu, Daniels et al., 2005). In the current dataset, there was no overall difference in response time for new items versus repeated items (1.49 s vs 1.53 s, \(p_{rep} = .75\)). This pattern suggests that participants, especially successful ones, did not follow a strategy of first assessing global familiarity (which would not differ for studied versus repeated words, but would for new versus repeated items) and subsequently attempt to identify the source. Instead, they may have evaluated the fit of each word presented at retrieval against the narrative structure (or other cognitive set) created at encoding. This
dimension differentiated studied versus unstudied words, but not new versus repeated words (c.f. Jacoby et al).

Source identification might be used to discriminate words that were familiar because they were on the studied list versus those that were repeated lures. Although there was no overall difference in response time for new versus repeated items, participants took longer to judge long-lag repeated lures than short-lag repeated lures (1.43 s vs 1.64 s, $p_{rep} = .99$). The longer time for long-lag items might reflect time spent trying to identify their source, and made a unique contribution even after accounting for PST. However, its relation to performance was negative, and did not eliminate age or ability effects.

A speculative interpretation of these patterns is that the optimal strategy is to proactively constrain retrieval to the studied list, and that advanced age and lower ability are associated with reduced self-initiation of the encoding processes that will support this strategy downstream. Separate from age and ability, when constrained retrieval fails, participants must engage in source identification. Greater difficulty with this process is associated with worse recollection.

In summary, we used principles from theories about self-initiation and proactive control to understand individual differences in training recollection. Self-initiation of control, here indexed by the relative amount of time spent on an intentional, open-ended encoding task, was a powerful predictor of individual differences in performance. Furthermore, it accounted for age and ability differences. Better encoding may have had its effects at least in part by supporting more proactive, source-constrained retrieval, allowing easier rejection of repeated lures. By contrast, a potential measure of retroactive attempts at source identification (time spent on long- versus short-lag repetitions), also predicted individual differences in performance, but did not account for those differences related to age and ability.

This paper focuses on the self-initiation of controlled processes at encoding, but similar principles may apply in any situation that creates strong demands for the self-initiation of control. For example, we predict that on a free recall task, the degree to which individuals spontaneously engage organizational strategies (e.g., semantic clustering) would be strongly related to age and ability differences in performance overall, and in the ability to benefit from recall training. Questions for future research include the link between putative measures of self-initiated processing and individual differences in executive function, working memory, and processing speed, and how the processes engaged at encoding may influence the tendency to engage in source-constraining or source-identifying processes at retrieval (see also work by Jacoby and colleagues).

More generally, we suggest that translating basic theory to the analysis of cognitive training programs holds great promise for understanding how those programs have their effects, and for improving them to benefit those most in need. In turn, observing which people improve, and how they improve, over the course of training can inform our understanding of the processes people apply to cognitive tasks.
References


Author Note

David Bissig, Wayne State University, Cindy Lustig, University of Michigan. We thank Janine Jennings for advice, Andrew Reinenke for assistance in preparing materials, and Megan Walsh, Lindsay Nelson, Mary Askren, and Elise Dagenbach for assistance with data collection.

Support provided by University of Michigan OVPR 5167.

Address correspondence to Cindy Lustig, University of Michigan, Ann Arbor, MI 48109-1043. Email clustig@umich.edu

Footnote

A potential concern is that advancement could represent only pre-existing recollection skills, rather than training-related improvement. We also tested 10 young adults (M age 20 yrs, SD = 1.3 yrs). For young adults, progression reflected initial ability: Almost all quickly achieved criterion at each lag level, usually with no more than one session at each level. (Uniformly high performance made it impossible to rank the young adults.) This was not the case for older adults: All but four remained at the lowest two lag levels for the first four study-test cycles. All conclusions remain the same if these four subjects are removed from analysis. No large differences for improvers versus nonimprovers were evident until the third day. These findings are consistent with those of Jennings et al. (2003, 2005), and with the idea that for older adults, improvement reflects training rather than initial ability.

Young adults had longer study times than did older adults (11.85 s vs 5.71 s) and shorter retrieval times (0.92 s vs 1.48 s), F(1,27) = 16.12, p_{rep} = .98. This cross-group comparison supports the idea that the emphasis on study versus retrieval contributes importantly to age differences in recollection. The study-time difference represents a rare instance of faster response times associated with worse performance and greater age.