Experimental Aging Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722867

Age Differences in the Use of Beneficial and Misleading Cues in Recall: With a Comment on the Measurement of Between-Group Differences in Accuracy

Aaron S. Benjamin[a]
[a] Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

Online publication date: 15 January 2011

To cite this Article Benjamin, Aaron S.(2011) 'Age Differences in the Use of Beneficial and Misleading Cues in Recall: With a Comment on the Measurement of Between-Group Differences in Accuracy', Experimental Aging Research, 37: 1, 63 — 75

To link to this Article: DOI: 10.1080/0361073X.2011.536742

URL: http://dx.doi.org/10.1080/0361073X.2011.536742

Experimental Aging Research, 37: 63–75, 2011

ISSN: 0361-073X print/1096-4657 online
DOI: 10.1080/0361073X.2011.536742

Routledge
Taylor & Francis Group

AGE DIFFERENCES IN THE USE OF BENEFICIAL AND MISLEADING CUES IN RECALL: WITH A COMMENT ON THE MEASUREMENT OF BETWEEN-GROUP DIFFERENCES IN ACCURACY

Aaron S. Benjamin

Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

Young and old subjects were tested on their memory for paired-associate terms when cued with either facilitative or misleading word stems. After studying a long list of pairs of unrelated words (e.g., hair–turtle), recall of a particular target term was cued in a facilitative manner (hair–tu____) or a misleading manner (hair–ta____). The effects of these cues were assessed relative to a baseline condition in which levels of performance lay between the other two (hair–t____). To interpret the age-related effects of the facilitative and misleading cues relative to baseline, the variance in differences between the baseline and the experimental conditions related to the overall baseline level was factored out, and age-related differences as a function of cue were assessed on the remaining variability. This analysis revealed that the two age groups differed both in their ability to overcome the adverse effects of the misleading cue and also to take advantage of the benefits afforded by the facilitative cue. This combination of results is consistent with the view that aging results in a loss of general strategic control, and not specifically inhibitory control, over the effects of retrieval cues.

A primary goal of current research in cognitive gerontology is the development of a parsimonious theoretical account of the effects of nonpathological aging on memory function. Not all memory skills

Received 4 April 2009; accepted 17 June 2009.
Address correspondence to Aaron S. Benjamin, Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820, USA. E-mail: ASBENJAM@UIUC.EDU

decline with age, and of those that do, some are more dramatic and debilitating than others. In this article, I pit two general contemporary classes of theories against one another in their ability to predict performance on a set of word fragment completion tasks. The first theory is that first proposed by Hasher and Zacks (1988; see also Hasher, Zacks, & May, 1999). In that framework, older adults suffer from a deficit of *inhibition*, whereby inappropriate, misleading, or extraneous thoughts are difficult to eliminate from consciousness. The second theory concerns the ability to deploy mnemonic control processes in a strategic manner, and is exemplified by such experiments as those reported by Hay and Jacoby (1999). They showed how the elderly suffer under conditions in which a habitual prepotent response needs to be suppressed in order to make a more controlled, nonhabitual response. This *control* view is not currently as precisely specified as the inhibitory view, which has made numerous predictions that have been borne out in experimental data (e.g., Hasher, Tonev, Lustig, & Zacks, 2001). Another interpretation of a lack of control is that the elderly are less apt to bring effective metacognition (Benjamin, 2003, 2005; Benjamin & Bawa, 2004; Castel, 2008) or retrieval strategies (Benjamin & Ross, 2008; Castel, Benjamin, Craik, & Watkins, 2002) to bear on the task.

For present purposes, the critical difference between the control view and the inhibition view is that a generalized lack of strategic mnemonic control in the elderly should cause both debilitated recall when a cue is misleading and an inability to take advantage of a recall cue that is facilitative. The inhibition view predicts that misleading cues should be detrimental for older adults, but—because inhibition presumably plays little role in the facilitative case—does not predict a difference between old and young in their ability to take advantage of facilitative cues. To illustrate how these theories play out in the context of the experiments reported here, I first discuss the general paradigm and then review these predictions in greater depth. For a more comprehensive review of these and other views of aging and memory, as well as empirical support, see Kester, Benjamin, Castel, and Craik (2002).

In these experiments, young and old subjects studied a series of paired-associate terms (e.g., *chair–water*) and were asked to learn them so that they would be able to recall the second term when provided with the first. During the recall test, subjects were then provided with one of several types of cues. In the baseline condition, they were given the cue word and the first letter of the target (*chair–w*____). In the facilitative condition, they were given the cue and the first two letters of the target (*chair–wa*___). Finally, in the misleading condition, they were provided the cue, the first letter of the target, and an incorrect second letter for the target (*chair–wo*___). All subjects were informed before the test that some of the target-stem recall cues would contain an incorrect letter, so the effects in this condition are not likely attributable to differences in "gullibility" between the groups.

It was assumed that mean performance in the facilitative condition would be superior to that in the baseline condition, and also that mean performance in the misleading condition would be poorer than baseline. The inhibition view clearly predicts that the elderly should suffer disproportionately, relative to the young, in the misleading condition, but does not naturally predict smaller benefits than the young from the provision of the facilitative cue. The control view predicts impairment in both conditions (and, indeed, in all conditions that emphasize the strategic component of retrieval): Relative to baseline performance, recall for the elderly should drop more in the misleading case and gain less in the facilitative case. The facilitative condition thus pits the control perspective against the inhibition view, and also against theoretical viewpoints (e.g., Craik, 1983) that predict that increasing "environmental support" will provide a relative advantage to the elderly.[1]

Measuring Accuracy on Bounded Scales

One methodological challenge in evaluating the hypotheses discussed above is that a substantial difference in recall baseline performance between the age groups is likely to obtain (cf. Craik & McDowd, 1987). Because there is no well-articulated mapping between the psychological construct of memory fidelity and the scale on which recall is measured, it is difficult to assess gains and losses relative to baseline. In general, higher baseline performance should lead to smaller gains and larger losses simply because of the scale range. This problem is exacerbated when evaluating means across heterogeneous groups. A secondary mission of this article is to illustrate a statistical manner of confronting this ubiquitous problem that permits the application of commonly used and well-understood linear analytic approaches.

The solution avoids the somewhat ungainly strategy of attempting to equate baseline performance by employing different manipulations

[1]Inconsistent with this prediction, Park and Shaw (1992) reported no difference between young and older subjects in the facilitation provided by additional letter cues, but the lack of an advantage for the younger group did not reconcile between the control and inhibition views.

of learning or study parameters between groups. That approach is problematic because, although the technique may functionally equate performance between groups, it does so at the expense of testing subjects on equivalent tasks. Thus, even though performance may be comparable, the underlying scale may have been altered in such a manner that the problem in interpreting between-group differences remains.

The question of how to best account for group differences in analytic comparisons has been the basis of much debate in research communities that employ response time (RT) as the primary dependent variable for interpreting age-related effects (Chapman, Chapman, Curran, & Miller, 1994; Faust, Balota, Spieler, & Ferraro, 1999; Myerson, Ferraro, Hale, & Lima, 1992), but the rigor brought to bear in that debate has not been widely imported into research addressing group differences in accuracy. That fact likely reflects in part the generally higher level of sophistication evident in the derivation of basic principles of response time analysis and comparison (e.g., Luce, 1986), but, at a general level, the problems faced when measuring accuracy are similar: that interactions between groups and experimentally manipulated variables—the basic stuff of theory-building in cognitive aging—may arise spuriously due to failures of the measurement scale to live up to the towering standards of linear analytic tools. However, whereas that failure in the analysis of RTs is principally due to the fact that effect magnitudes scale with baseline scores (e.g., Salthouse, 1985), accuracy suffers because it is typically measured on scales that are bounded on both ends (such as proportions).[2] Scale boundaries compromise the validity of traditional analytic approaches because the asymptotic distribution of the central limit theorem is symmetric and defined over an infinite range. That is, analyses that assume linearity of the measurement scale are doomed to fail when values—like proportions—are restricted. This problem is particularly apparent when scale boundaries are approached and the tails of the distribution are differentially truncated.

Here I address this problem with a combination of methodological and statistical techniques. First, I employ a manipulation of learning to ensure that performance on the baseline variable overlaps across the age groups. The reasons for this will shortly be apparent. Second,

[2]Distance-based measures of accuracy (such as the commonly used d') have an advantage over proportions, in that they are only bounded on the low end of the scale (see, e.g., Benjamin, Diaz, & Wee, 2009; Matzen & Benjamin, 2009).

and consistent with the typical approach in psychology, I examine not raw performance in the experimental conditions, but rather differences between performance in those conditions and the baseline condition. Although this rudimentary correction eliminates some of the variability between facilitative and misleading conditions that is systematically related to baseline performance, it assumes that the contributions of those conditions sum linearly across the measurement scale (again: an unlikely possibility, given the bounded nature of proportional scales). Thus, I also factor out of these difference scores the linear relation between baseline performance and those difference variables assessed over *both* age groups. The residual scores reflect the degree to which the benefits of the facilitative cue (or the costs of the misleading cue) are greater or lesser than those linearly predicted by baseline performance. Because any scalar effects are, by definition, *not* group specific, those effects are best estimated by ignoring group membership. It is for this reason that it is critical to have overlapping performance ranges between the groups: otherwise, it is possible that the regression lines relating scale location to effect magnitude are disproportionately influenced by group differences and thus poorly reflect the scalar effects in question (as in, for example, Simpson's Paradox [Simpson, 1951]).

The residual effects can then be tested *across* age groups, and should allow us to claim with greater confidence that the magnitude of the experimental effects will be relatively free of scaling artifacts. This technique assumes that the scaling effects (as measured by the regression) and the effects of the experimental manipulation sum additively, and is thus subject to the same concerns about the linearity of the underlying effects, but—because it accounts for at least a proportion of the purely scalar sources of variance—is less subject to those worries than are most statistical applications of linear models.

EXPERIMENT

In this experiment, we assessed the ability of younger and older adults to complete word stems with previously studied words when provided with the three types of cues described above. To reiterate, the loss of control view of aging predicts greater losses with misleading cues and smaller benefits with facilitative cues for the elderly, whereas the inhibition view predicts only greater losses, but not lessened benefits. Two different study times were employed in order to ensure a greater amount of overlap between the age groups.

METHODS

Subjects

Forty-two young (mean age = 20, females = 34) and 37 elderly subjects (mean age = 71, females = 23) participated in Experiment 1. The young subjects were undergraduates in an introductory Psychology course at a large public university, and the elderly subjects were healthy volunteers recruited through the media. Performance on the Mill Hill test of vocabulary was superior for the elderly (73%) than for the young (49%).

Design

The experiment used a 2 (age) × 3 (cue type) × 2 (study duration) mixed factorial design. Cue type was manipulated completely within subjects. Study time was manipulated between subjects.

Materials

Fifty-seven word pairs were used. The words in each pair were chosen to have little or no semantic relation to one another, ranged from three to seven letters, and were medium-frequency nouns, adjectives, and verbs. Each subject studied the same 57 pairs, but in a different randomly generated order.

For the test, each target word was assigned three potential test cues, one of which was simply the first letter, followed by the appropriate number of dashes (p—for pear), one of which was the first two letters (pe—), and the last of which was two letters, the second of which was incorrect (pi–). The misleading second letter was always chosen to allow multiple completions of the word, although no attempt was made to equate the number of potential completions across the conditions. The order of items on the test was also randomly generated for each subject. One third of the pairs were assigned to each of the three cue conditions.

Procedure

Prior to the study, all subjects were informed that they would be viewing pairs of words and instructed to learn the pair such that they could recall the second word when given the first. If there were no questions, subjects cycled through all 57 items at a rate of 3 or 6 seconds/item with a 1-s interitem interval. After study and a ∼1-min distractor phase during which subjects answered brief autobiographical questions, instructions for the test were given. Subjects were told that they would be provided with the first (cue) word and part of the second (target) word on the computer screen and should type in the entire target word, if they could remember it. They were also informed that they would sometimes be given just the first letter and sometimes the first two letters of the target word, but that second letter might be wrong. An example of all three conditions was provided in the instructions. During the test, subjects entered the target word when they could remember it, and skipped the trial or guessed when they could not. Errors of commission and omission were scored the same. After the experiment, all subjects were debriefed and shown their results.

RESULTS

The raw data are shown in Figure 1. Because the study time variable was not of relevance to the predictions tested here, and because it did

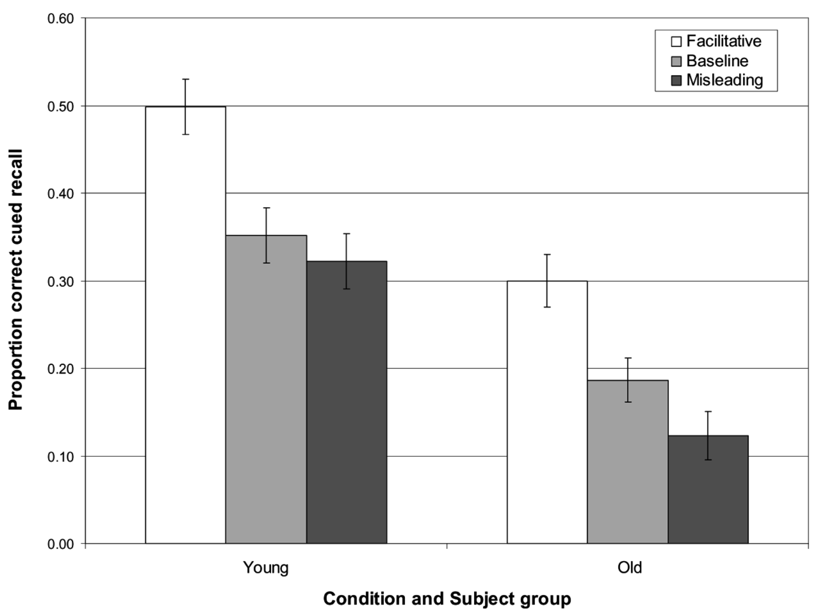

Figure 1. Proportion of correct responses to facilitative, baseline, and misleading items as a function of age. Error bars show the standard error of the mean.

not interact with the other variables, it is collapsed across in Figure 1. The results of all inferential tests reported here and throughout the paper are reliable at the $\alpha = .05$ level using two-tailed tests unless otherwise noted. Separate-variance *t* tests are employed for all between-group comparisons using degrees of freedom approximated via Welch-Aspin estimation (Aspin, 1948; Welch, 1947).

Examination of baseline performance revealed differences between the age groups, as expected (separate-variance $t(74.6) = 4.08$). At this level of analysis, young (paired $t(41) = 5.96$) and old (paired $t(35) = 6.26$) subjects showed a beneficial effect of the facilitative cue. In addition, the detrimental effect of the misleading cue reached significance for the performance of old (paired $t(35) = 3.77$) but not young (paired $t(41) = 1.47$) subjects. As noted before, this analysis is naturally subject to serious concerns about the nature of the measurement scale. More importantly, comparing performance between the age groups in the two nonbaseline conditions—the critical test of the hypotheses laid out here—is even more awkward because of the between-group differences in mean baseline performance. Thus, the three scores were converted to differences from

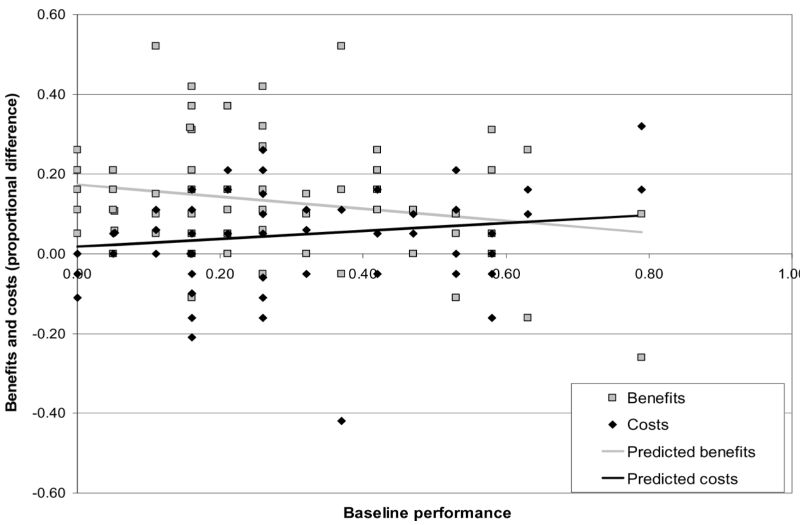

Figure 2. Benefits of the facilitative cue and costs of the misleading cue, relative to baseline, as a function of baseline level. Regressions show a likely artifactual age-independent relation between baseline and the two derived measures.

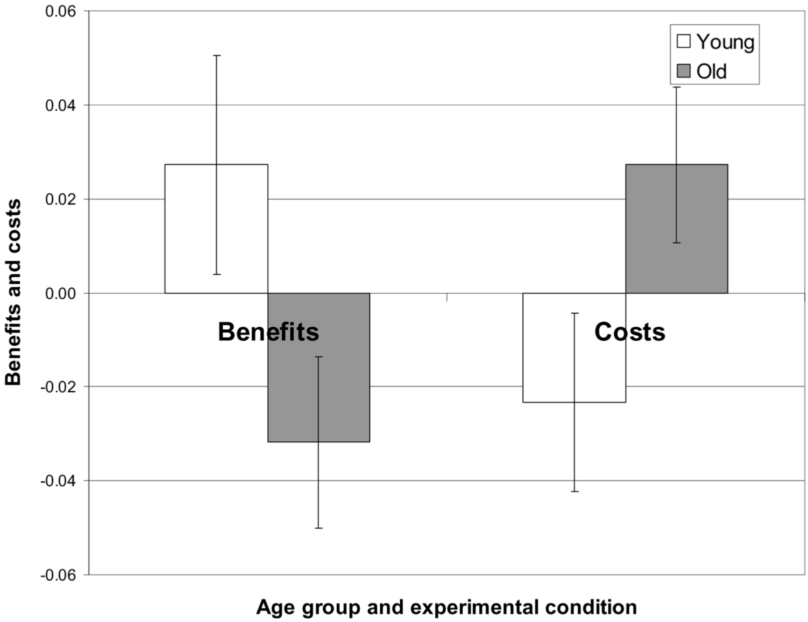

Figure 3. Residual benefits and costs as a function of age group following removal of the effect shown in Figure 2. Positive benefits indicate superior performance; positive costs indicate poorer performance. Error bars show the standard error of the mean.

baseline. Neither benefits (separate-variance $t(72.3) = 1.12$) nor costs (separate-variance $t(75.5) = 1.32$) differed reliably between age groups on these corrected scores.

Although this correction serves to remove variance in performance that owe strictly to differences in baseline between the age groups; it does *not* address the important confound that those differences in baseline likely induce differential scalar artifacts across the two within-subject conditions. To address this problem, the difference scores (benefits and costs) were regressed onto baseline performance *across* the age groups, and the residual benefits and costs were analyzed. The regressions are shown in Figure 2 and illustrate the hypothesized effect: As baseline performance increased, benefits became smaller and costs became larger. Figure 3 shows the residual effects and reveals a reliable difference between the age groups for both costs (separate-variance $t(75.7) = 2.01$) and benefits (separate-variance $t(74) = 2.00$).

DISCUSSION

It is interesting to note that the interpretation of the effects in this experiment depend critically on assumptions made about the nature of measurement scales. The raw data revealed one pattern, the deviations yet another, and the analysis of residuals told a third story. Psychologists often trust their eyes more than their theories, and one might be tempted to argue that the patterns evident in Figure 1—in which age differences arose in the misleading but not facilitative condition—are "purer" than their mongrel cousins shown in Figure 3. I do not wish to proselytize about the virtues of the present analytic approach, but I do want to emphasize that an acknowledgement of underlying assumptions is always in order. The assumptions made here are straightforward but not entirely noncontentious: that scaling effects on difference scores sum linearly with true condition effects to yield the obtained measurements. To rely upon the untransformed data, one must be willing to accept that the apparata of the general linear model is appropriate for the current measurement scale. I believe that scale boundaries seriously compromise that application, and have thus described one way to attempt to tease out a portion of the scale-related effects (see also Faust et al., 1992).

The results of our experiments reveal that the elderly suffer more under cuing conditions that are misleading, a finding that could be expected on the basis of the wide array of data showing an age-related deficit in inhibiting irrelevant information that comes to mind. In the misleading cuing condition, the invalid cue elicits numerous completions, none of which are the correct response. Young subjects may be more able to successfully inhibit the conscious products of this cue when no answer is apparent, and perhaps engage the same retrieval processes that they use when only the first letter of the target word is available, as in the baseline condition. A plausible explanation for the inability of older subjects is that they are unable to suppress the automatic influences of the cue, and cannot keep the retrieval products of that cue out of working memory long enough to engage in an undisrupted retrieval attempt using only the first letter. This phenomenon is perhaps analogous to the situation in which an answer to a particular crossword clue feels more difficult to access because of misleading letters garnered from incorrect neighboring responses. It is thus conceivably advantageous to maintain a fairly conservative criterion for responding in such puzzles, lest one be misled by previous incorrect answers!

However, it is also the case that older subjects benefit less than their younger counterparts from the provision of an additional facilitative retrieval cue. Presumably, unlike the case with the invalid cue, this difficulty does not arise from an inability to suppress irrelevant and misleading responses that come to mind. It does appear that the beneficial cue either orthographically constrains responses that come to mind more effectively or serves as a useful external validation for the responses that do come to mind, because recall performance does improve relative to the baseline condition for both young and old subjects. The age-related disadvantage in this condition (with facilitative cues) is not naturally handled by the simple version of the deficient-inhibition view of age-related deficits in memory presented here.

It could, however, be the case that the same problem of inhibition—namely, suppressing completion alternatives—impairs the performance of older subjects in the facilitative as well as the misleading condition. Whereas the misleading condition leads to the generation of candidates, none of which have been studied, the facilitative condition leads to the generation of candidates, only one of which has been studied. If that number of candidates is high, then the necessity to suppress alternatives is nontrivial, even in the facilitative condition. Thus, an extension of the deficient-inhibition view can handle age-related impairments in the facilitative condition, but additionally predicts that the effects of age should be smaller than the misleading condition (because the correct answer is more likely to be a part of the set of generated possibilities in the facilitative than misleading condition). As can be seen in Figure 3, that effect does not obtain—in fact, the magnitude of the age-related effect is somewhat larger in the facilitative condition.

The obtained results are, however, consistent with the broad view that older adults exert generally less control over strategic aspects of retrieval situations, such as accurately perceiving external cues, translating those cues into useful internal search cues, and filtering the retrieved products effectively (e.g., Benjamin, 2008; Castel, 2008). The young adults in our experiment might adopt a strategy, for example, of searching first for words that match the first letter of the cue—regardless of whether a second is presented—and then only using the second letter if the first retrieval attempt is unsuccessful. If older subjects simply use the strategy of submitting to memory whatever retrieval cues are available, even if they are known to be of occasionally limited potential, then they will have compromised their retrieval in two ways. First, the misleading cues will lead them astray and decrease their chances for eventually arriving at the answer. Second, they have limited the number of useful interrogative cues in the facilitative case to one, whereas the young have two. Although it is obvious that the world of possible completions to "st—" must be a logical subset of those to "s——," it is not necessarily the case that the retrieval products of the two cues share the same nested relationship, nor that the strategic search initiated by each employs the same parameters (cf. Tversky & Kahneman, 1983).

This explanation is entirely speculative, of course, but it does provide an example of a parsimonious view whereby both effects evident in our data can be explained with a single putative mechanism. It is my view that any comprehensive view of memory and aging must account for a general loss of strategic control over the use of memory, rather than a specific loss of inhibitory processes.

REFERENCES

Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika, 35,* 88–96.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31,* 297–305.

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

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

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory & Language, 51,* 159–172.

Benjamin, A. S., Diaz, M. L., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review, 116,* 84–115.

Benjamin, A. S., & Ross, B. H. (2008). *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48). London: Academic Press.

Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 225–270). London: Academic Press.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, Series B, 302,* 341–359.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13,* 474–479.

Chapman, L. J., Chapman, J. P., Curran, T. E., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question. *Developmental Review, 14,* 159–185.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125,* 777–799.

Hasher, L., Tonev, S. T., Lustig, C., & Zacks, R. T. (2001). Inhibitory control, environmental support, and self-initiated processing in aging. In M. Naveh-Benjamin, M. Moscovitch, & R. L. Roediger, III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honour of Fergus Craik* (pp. 286–297). East Sussex, UK: Psychology Press.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press, Inc.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: The MIT Press.

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. *Psychology & Aging, 14,* 122–134.

Kester, J. D., Benjamin, A. S., Castel, A. D., & Craik, F. I. M. (2002). Memory in the elderly. In A. Baddeley, B. A. Wilson, & M. D. Kopelman (Eds.), *Handbook of memory disorders* (2nd ed., pp. 543–567). Chichester, UK: Wiley.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Matzen, L. E., & Benjamin, A. S. (2009). Remembering words not presented in sentences: How study context changes patterns of false memories. *Memory & Cognition, 37,* 52–64.

Myerson, J., Ferraro, F. R., Hale, S., & Lima, S. D. (1992). General slowing in semantic priming and word recognition. *Psychology and Aging, 7,* 257–270.

Park, D. C., & Shaw, R. J. (1992). Effect of environmental support on implicit and explicit memory in younger and older adults. *Psychology and Aging, 7,* 632–642.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 400–426). New York: Van Nostrand Reinhold.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B, 13,* 238–241.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika, 34,* 28–35.