Paivio, A. (2007). *Mind and its evolution: A dual-coding theoretical approach.* Mahwah, NJ: Erlbaum.
Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475–543.
Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1139–1153.
Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger, III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.
Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 77–90). Mahwah, NJ: Erlbaum.
Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46,* 419–440.
Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E., Jr. (1995). The effects of contextual interference on the acquisition and retention of logical rules. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 95–131). Thousand Oaks, CA: Sage.
Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 179–187.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.
Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2007). Pushing the limits of imagination: Mental practice for learning sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 254–261.
Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008). A mental practice superiority effect: Less retroactive interference and more transfer than physical practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 823–833.

# 15

# AN OUTPUT-BOUND PERSPECTIVE ON FALSE MEMORIES

## *The Case of the Deese–Roediger–McDermott (DRM) Paradigm*

Asher Koriat, Ainat Pansky, and Morris Goldsmith

Recent years have seen an upsurge of interest in memory accuracy and distortion (Koriat, Goldsmith, & Pansky, 2000). This interest has been fueled by a host of real-life observations documenting severe memory distortions and fabrications, casting doubt on the faithfulness of eyewitness memory (Loftus, 1979, 2003). Some of the studies on memory distortion and false memories have examined naturally occurring memory errors that derive from the constructive nature of memory and are in line with the view originally advanced by Bartlett (1932). Other research has shown how memory is sensitive to a variety of influences that result in erroneous memories (for a review, see Pansky, Koriat, & Goldsmith, 2005). All in all, the view of memory that seems to emerge from the research literature is rather pessimistic regarding the ability of memory to deliver a veracious account of past events. This view is reflected, for example, in the title of Schacter's (2001) book, *The Seven Sins of Memory.*

In this article we focus on the phenomenon of false recall, and in particular on what is perhaps the most impressive laboratory manifestation of this phenomenon, documented by Roediger and McDermott (1995) and widely replicated since. In the Deese–Roediger–McDermott

(DRM) paradigm, a study list is presented, composed of words (e.g., *thread, pin, eye, sew*) that are associates of a critical nonpresented word (e.g., *needle*). Participants are found to exhibit high rates of false recall of the critical lure even when they are urged not to guess. The DRM paradigm has yielded a wealth of findings on false recall and has provided important insights about the processes underlying memory errors, the factors that affect the rate of these errors, and the extent to which such errors can be avoided (see Roediger, McDermott, & Robinson, 1998, for a review).

In this chapter we address a seemingly simple question: What general message should the DRM research deliver to the scientific community and to the general public regarding the reliability of human memory? What do the findings tell us about the extent to which memory reports about past events can be trusted? It might be argued that the stimulus situation used in the DRM paradigm—the study of a list of 15 or so words that are all related to a single word that is itself absent—is not ecologically representative, and perhaps for that reason the implications of the DRM results for everyday memory are limited (e.g., Freyd & Gleaves, 1996). We shall put this argument aside for now, and examine the message that follows from the findings on the assumption that the DRM conditions are in fact representative of real-life memory situations. We should stress that until now, the focus and target of DRM research has been to clarify the mechanisms by which false memories may be created or avoided, rather than to convey a general message concerning the dependability of memory reports as a whole. Nevertheless, such a message seems to emanate—at least implicitly—from DRM findings.

Consider the basic observation in DRM studies. The rate of false memory, measured by the probability of reporting the critical nonpresented word, is quite startling: On immediate testing, it is about the same as that of recalling studied words from the middle of the list (assumed to reflect retrieval from long-term memory; Roediger & McDermott, 1995; Schacter, Verfaellie, & Pradere, 1996). On delayed tests, it tends to be even higher than that of studied items (McDermott, 1996). What is more, false recalls are remarkably persistent over time: Whereas the proportion of correctly recalled items reveals the typical decline with retention interval, the probability of recalling the nonpresented item tends to remain high or even to increase (McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Seamon, Lao, Kopecky et al., 2002a; Toglia, Neuschatz, & Goodwin, 1999). Also, whereas veridical recall tends to remain stable across repeated testing (following a single study presentation), false recalls tend to increase (Payne et al., 1996). Overall, it would appear that false memories in the DRM paradigm are no less frequent, and even more persistent, than true memories (see Roediger et al., 1998).



What conclusion is a layperson or a judge to draw from these findings regarding the overall trustworthiness of memory reports? If witnesses are as likely to falsely remember a nonpresented item as they are to correctly remember a presented item, would it not seem natural to conclude that memory reports are worthless? To consider this question, we must first clarify the distinction between input-bound and output-bound memory assessment.

## INPUT-BOUND AND OUTPUT-BOUND MEASURES OF MEMORY PERFORMANCE

Traditionally, measures of memory have been calculated conditional on the input, by expressing the number of items recalled or recognized as the proportion or percentage of the total number of items presented. The assessment of memory performance in terms of input-bound percent correct follows naturally from the storehouse metaphor that underlies much of traditional memory research (Koriat & Goldsmith, 1996a; Roediger, 1980). Koriat and Goldsmith (1994, 1996a, 1996b; for a review, see Goldsmith & Koriat, 2008) have referred to such measures of memory performance as *quantity* measures, because they are assumed to reflect the amount of presented or studied information that has been retained and is currently accessible.

Memory performance, however, can also be assessed using *output-bound* measures, in which the number of correct items recalled is expressed as a proportion or percentage of the total number of items *reported*. Such measures reflect the *accuracy* of the memory report, in terms of the probability that a reported item is correct. Consider, for example, a participant who is presented with 50 words, and in a free-recall test reports 40 words, 36 of which are correct and 4 are commission errors. Input-bound memory quantity performance in that case is .72 (36/50); that is, 72% of the input-study items have been successfully recalled. In contrast, output-bound memory accuracy is .90 (36/40). That is, 90% of the output-recalled items are, in fact, correct. This latter measure uniquely reflects the *dependability* of the information that is reported—the degree to which each reported item can be trusted to be correct.

It is important to stress that output-bound accuracy and input-bound quantity measures can be distinguished operationally only when participants are given the option of *free report*. On forced-report tests, such as forced-choice recognition or (less commonly) forced recall, participants are required to provide a substantive response to each and every test item; "pass" or "don't know" responses are not allowed. Under such conditions, the input-bound quantity and output-bound accuracy percentages are necessarily equivalent, because the number of output items is the same as the number of input items (see Koriat & Goldsmith, 1994, 1996a). For example, if a participant gets 40 out of 50 choices correct on a forced-choice recognition test, we may conclude either that the probability of correctly recognizing an input item is .80 (input-bound quantity) or that the probability that a reported item is correct is .80 (output-bound accuracy). The difference between the two measures is entirely a matter of interpretation—whether one intends to measure quantity or accuracy. In contrast, on free-report tests, such as cued or free recall, participants are allowed to omit items from the memory report or, equivalently, to respond "don't know" if they feel they do not remember an item. In this case, the number of output items may be far fewer than the number of input items. In this chapter we consider only DRM results obtained under free-report conditions, in which input-bound and output-bound measures differ operationally as well as conceptually.

Although the focus of false memory research is on memory accuracy, the analyses of false recall performance in the DRM paradigm generally follow the logic underlying the computation of input-bound performance: They focus on the probability of recalling the critical, nonpresented item under various conditions, and compare this to the probability of recalling a studied list item. This focus perhaps reflects a treatment of the critical lure as if it were an implicit study item (see activation accounts of DRM performance; e.g., Roediger, Balota, & Watson, 2001; Roediger & McDermott, 2000; Roediger, Watson, McDermott, & Gallo, 2001). However, for an external observer, such as a courtroom judge, who is concerned by the phenomenon of false memory, the output-bound accuracy measure is arguably of greatest concern: To what extent can we depend on what a witness reports to be true? That is, what is the probability that an item of information reported by a witness is correct? If the witness reports that there was a knife at the scene of the crime, what is the likelihood that indeed a knife was present? This is the conditional probability captured by output-bound accuracy.

What do we know about output-bound memory accuracy in general? A cursory examination of the literature suggests that the accuracy of what people report under free-report conditions is high—typically in the range between .80 and .95, even following long retention intervals (Ebbesen & Rienick, 1998; Koriat, 1993; Koriat & Goldsmith, 1994, 1996b; Poole & White, 1993). This impressive level derives largely from



metacognitive monitoring and control processes that operate under free-report conditions. A number of studies (e.g., Kelley & Sahakyan, 2003; Koriat & Goldsmith, 1994, 1996b) have shown that when given the option to choose which items to report and which to withhold, people enhance their memory accuracy considerably in comparison to forced-report testing, and do so by screening out answers that are likely to be wrong. Koriat and Goldsmith (1996b) proposed a model of the strategic regulation of memory reporting in which rememberers monitor the likelihood that each candidate memory response is correct, and then compare that likelihood to a preset report criterion to determine whether to volunteer that response. Because the control decision is based on the subjective confidence associated with each item that comes to mind, and confidence is generally predictive of correctness, participants are generally effective in regulating their reporting so as to enhance accuracy when accuracy is at stake. Thus, for example, Koriat and Goldsmith (1994, Experiment 1) found that simply giving participants the option of free report allowed them to increase their output-bound accuracy substantially compared to forced-report accuracy, and giving them a stronger incentive led them to increase accuracy even further (Koriat & Goldsmith, 1996b, Experiment 3). In fact, in the latter experiment, fully 25% of the participants were successful in achieving 100% accuracy (see also Higham, 2007; Kelley & Sahakyan, 2003; Koriat & Goldsmith, 1996b)! Similarly high levels of accuracy under free-report conditions have been observed in children as young as eight years old (Koriat, Goldsmith, Schneider, & Nakash-Dura, 2001; Roebers & Fernandez, 2002).

To what extent can people draw an output-bound conclusion on the basis of data from an input-bound performance? We examined this question with regard to DRM findings in an informal study conducted with Haifa University third-year psychology undergraduates who were enrolled in a research seminar on memory distortions. The students ($n = 38$) read the chapter by Roediger and Gallo (2004) summarizing DRM findings and theories, discussed the chapter (in groups of 12 or 13), and were finally asked to answer several multiple choice questions about the DRM phenomenon. Among the questions were two key ones that asked about the implications of the DRM findings from an output-bound perspective. The questions and the distribution of responses appear in Table 15.1. It can be seen that the correct answers to questions 1 and 2 were selected in only 8 and 11% of the cases, respectively. This informal example serves to illustrate the idea that people have difficulty shifting from an explicitly presented input-bound perspective to an output-bound perspective.

**Table 15.1** The Relative Frequency in Which Each Answer Was Chosen on the Two Key Questions in Our Survey ($n = 38$) (The Correct Answer Is Highlighted in Bold)

| Question | Optional Answers | Relative Frequency |
|---|---|---|
| 1. In the DRM paradigm, when a participant recalls a particular word, its chances of being a critical lure that had not appeared in the study list are: | a. Higher than its chances of being a studied word | 8% |
| | b. Lower than its chances of being a studied word | 8% |
| | c. More or less equal to its chances of being a studied word | 84% |
| 2. It is possible that the DRM situation is not representative of everyday situations. However, assume that there are situations in real life that are very similar to the one experienced by a participant in a DRM experiment, and that a person in such a situation must later testify in a courtroom. In your opinion, which of the following statements best describes the practical implication of the studies conducted using the DRM paradigm with regard to the extent to which a judge (or juror) can rely on this person's testimony? | a. To the extent that a real-life situation resembles the DRM situation, a judge/juror can rely on most of the information that an eyewitness provides as being correct. | 11% |
| | b. To the extent that a real-life situation resembles the DRM situation, a judge/juror cannot rely on most of the information that an eyewitness provides as being correct. | 47% |
| | c. To the extent that a real-life situation resembles the DRM situation, a judge/juror can rely on about half of the information that an eyewitness provides as being correct. | 37% |
| | d. To the extent that a real-life situation resembles the DRM situation, the eyewitness report is useless. | 5% |



## OUTPUT-BOUND MEMORY ACCURACY IN STUDIES USING THE DRM PARADIGM

We now review previous DRM studies with a focus on the output-bound accuracy of recall.[1] As mentioned earlier, output-bound memory performance is calculated as the proportion of correctly recalled study items out of the total number of items recalled (i.e., reported). The number of correctly recalled study items is routinely reported in all published DRM studies or can be readily calculated. The total number of items recalled is sometimes reported, but if not, it too can be calculated by summing the number of correctly reported items and the number of commission errors. A remaining problem, however, is that although DRM articles always report the rate of a particular commission error—the so-called *critical lure* (e.g., *sleep* in a list of sleep-related words)—they often fail to provide information about *other* commission errors (e.g., *dream*, *pillow*, or other sleep-related or unrelated nonpresented words). Estimates of output-bound accuracy performance that do not take these noncritical commission errors into account will be inflated. Therefore, in the analysis of DRM results, presented below, we included only studies in which data are reported for both types of commission errors.

Based on the information just mentioned, output-bound accuracy (*OBA*) was calculated for each study (or experimental condition) as follows:

$$OBA = \frac{Ps \cdot Ni \cdot Nl}{Nr} = \frac{Ps \cdot Ni \cdot Nl}{(Ps \cdot Ni \cdot Nl) + (Pc \cdot Nl) + (Nnc \cdot Nl)} \qquad (15.1)$$

where:

$Nl =$ Total number of lists
$Ni =$ Number of items in each input list
$Nr =$ Total number of items recalled
$Ps =$ Probability of recalling the studied items
$Pc =$ Probability of recalling the critical lure
$Nnc =$ Number of noncritical commission errors

Our analysis was based on a total of 108 published DRM studies that allowed the calculation of output-bound recall accuracy by (1) allowing participants the option of free report, and (2) reporting all of the relevant data (including *Nnc*). Many of these studies consisted of different experimental conditions in the same research report.[2] Table 15.2 (see appendix at the end of this chapter) presents a list of these studies, a brief description of the conditions used in each study, and the main data that entered into the calculation of *OBA*. Not presented in the table, the mean number of noncritical commission errors (*Nnc*) reported by each participant in each study was 2.7 (5% of all reported items), whereas the mean number of critical lure errors was 3.2 (6% of all reported items). Thus, the rate of producing the single critical lure item was somewhat higher than that of all other commission errors combined.

Figure 15.1 presents the distribution of *OBA* scores across the 108 studies. One result is particularly striking: Output-bound accuracy exceeded .90 in the majority of studies, and it was rarely lower than .85. In fact, mean *OBA* was .89 (SD = 0.10) and the median was .92. We also calculated mean *weighted OBA* by weighting each study mean by the number of participants on which it was based. This mean was also .89 (SD = 0.09). Note that these values were obtained across all studies, including some that used children, older adults, and amnesic patients, as will be discussed below. If we limit our analysis to the "standard" DRM conditions, such as those originally used by Roediger and McDermott (1995)—examining young adults' immediate recall of each list following intentional encoding (N = 69)—an item recalled from a DRM list has a mean likelihood of over .93 of being correct.

We compared these values to two input-bound measures. The first is the *input-bound quantity* (*IBQ*) score—the probability of recalling a studied item (equivalent to *Ps* in Equation 15.1). Mean *IBQ* was .50 (SD = 0.18), and median *IBQ* was .55. The *weighted IBQ* mean was .52

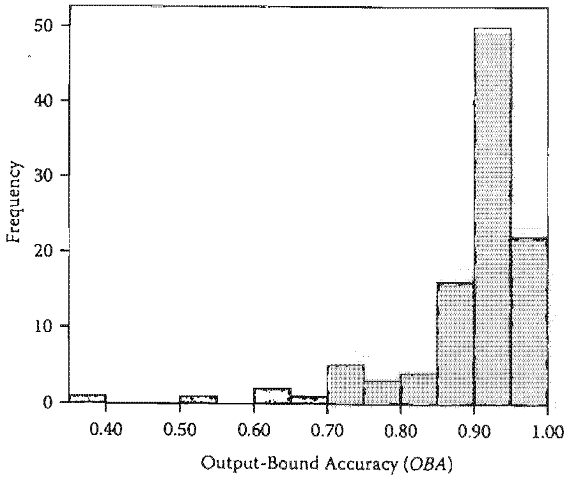


**Figure 15.1** Frequency distribution of mean output-bound accuracy (*OBA*) performance across the 108 studies.



(SD = 0.18). Thus, as expected, *OBA* was considerably higher than *IBQ*. That is, even though participants may not remember much of the input information, the vast majority of what they do report is correct.

The second measure, *Pc*, is the probability of recalling a particular nonstudied item—the critical lure. Although not a true input-bound index, as noted earlier, *Pc* is based on the same underlying logic, which is most easily seen if one treats the critical lure as an implicit study item. This is the measure that has been the focus of most DRM studies. Mean *Pc* across the 108 studies was .34 (SD = 0.15), the median *Pc* was also .34, and the *weighted Pc* mean was .35 (SD = 0.15). Thus, across these studies, the likelihood of falsely recalling the critical lure was somewhat lower than that of correctly recalling a studied item, but was still quite high.

What is the relationship between the *OBA* and *Pc* measures? One might expect these two measures to be inversely related, with the former indexing accuracy and the latter indexing error rate. However, the correlation between the two measures across the DRM studies was, in fact, virtually zero ($r = -.02$). Inspection of the bivariate distribution in Figure 15.2 suggests that indeed, for the majority of studies ($n = 73$), those in which *OBA* was .90 or more, the correlation is negative ($r = -.58$), as would be expected. In contrast, for the relatively small number of studies in which *OBA* was below .90, the correlation is positive ($r = .47$, $n = 32$); that is, a higher rate of critical lure intrusions tends to be

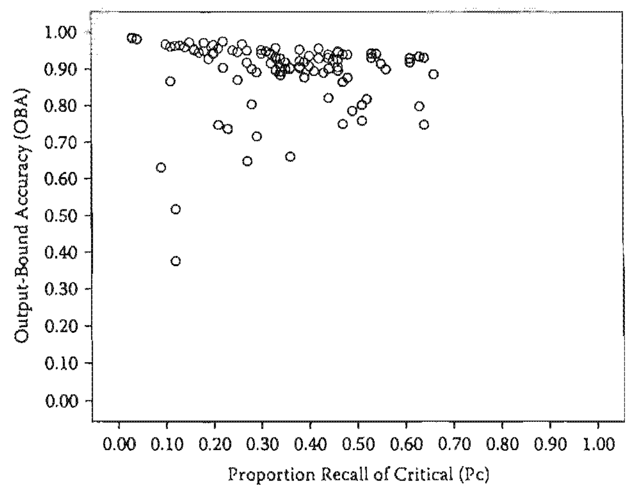


**Figure 15.2** The bivariate distribution of the proportion of recall of the critical nonpresented item (*Pc*) and the mean output-bound accuracy (*OBA*) performance scores across the 108 studies.

associated with *better* output-bound accuracy. This latter correlation may be mediated by differences in *IBQ*, as will be discussed below.

The comparison between *OBA* and *Pc* illustrates the contrast between two different perspectives for the examination of DRM results. The first, which is characteristic of all DRM studies, is an input-bound perspective that focuses on the rate at which a predesignated, critical lure is reported. The second is an output-bound perspective, which focuses on the likelihood that each reported item is correct. It is the latter perspective that should be of interest to a courtroom judge or to any external observer who is concerned with the dependability of the memory report as a whole. Thus, it is important to note that despite the remarkable success of the DRM paradigm in inducing a substantial rate of false recall, as reflected in a mean *Pc* of .34, the overall output-bound accuracy (dependability) of the participants' memory reports is nevertheless very high, with more than two-thirds of the sampled studies yielding an *OBA* performance of .90 or higher.

A detailed examination of the results presented in Table 15.2 highlights some interesting trends in *OBA* across different populations and conditions.[3] First, as shown in Figure 15.3, young adults exhibited higher accuracy than older adults, children (aged 10 or less), and amnesic patients.[4] As shown in Figure 15.4a, using longer retention intervals

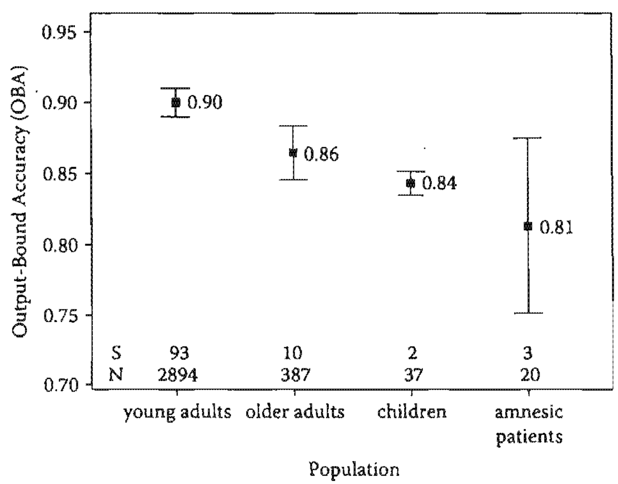


**Figure 15.3** Mean output-bound accuracy (*OBA*) performance for studies using young adults, older adults, children, and amnesic patients. S = number of studies, N = number of participants across these studies. Error bars represent ± 1 SEM.



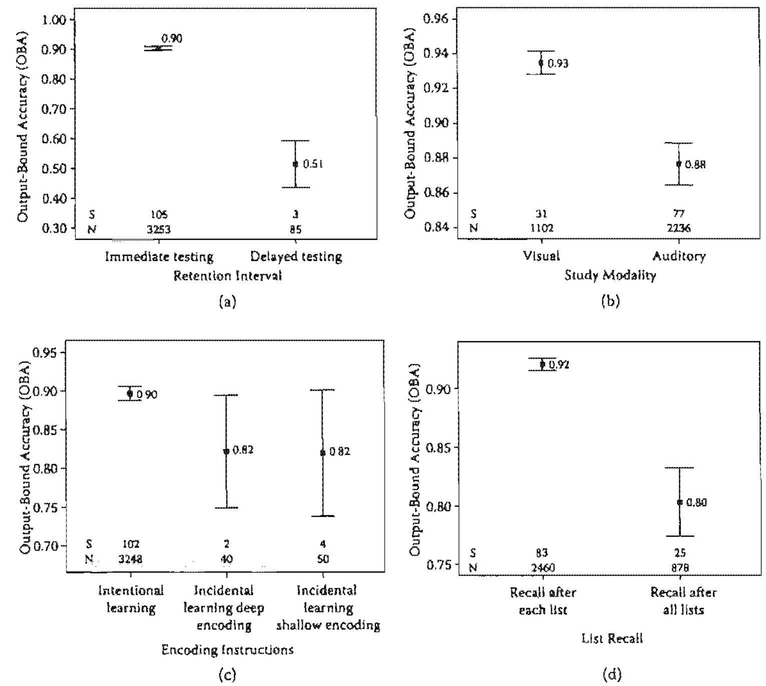


**Figure 15.4** (a) Mean output-bound accuracy (*OBA*) performance for studies using immediate testing versus those using delayed testing. (b) Visual presentation versus auditory presentation. (c) Intentional learning versus incidental learning (deep or shallow encoding). (d) Studies in which recall was measured after each list versus those in which it was assessed after the study of all lists (Figure 15.5b). S = number of studies, N = number of participants across these studies. Error bars represent ± 1 SEM.

between study and test (ranging from 48 hours to 2 months) reduced *OBA* considerably, compared to immediate testing (cf. Goldsmith, Koriat, & Pansky, 2005). Additional factors that seem to influence *OBA* are presentation modality (visual vs. auditory; see Figure 15.4b), encoding instructions (intentional vs. incidental learning at different levels of processing; see Figure 15.4c), and whether each list is tested individually following its presentation or in a single joint test following the presentation of all the lists (see Figure 15.4d). These results are suggestive of the many factors that affect the dependability of memory reports, and which deserve to be studied more systematically with a focus on output-bound memory performance (see Koriat et al., 2000, for discussion).

As might be expected, the ordering of the different populations in terms of *OBA* performance corresponds roughly with their ordering in terms of *IBQ* performance. Thus, for example, *IBQ* performance for young adults, older adults, children, and amnesic patients was .51, .47, .42, and .28, respectively. A similar pattern exists for the results presented in Figure 15.4a to d. In fact, as seen in Figure 15.5, *OBA* and *IBQ* are strongly and positively correlated across the DRM studies ($r = .72$). This is to be expected because of the mathematical relationship between the two variables (both have the same numerator—number of correct reported answers). In addition, however, this correlation may reflect a positive relationship between memory retention and memory monitoring. Indeed, several factors that improve retention (e.g., increasing the number of presentations of the study lists, increasing item presentation duration, allowing full vs. divided attention at study) have been found to improve memory monitoring as well (e.g., Benjamin, 2001; Kelley & Sahakyan, 2003; McDermott & Watson, 2001).

Returning now to the relationship between *OBA* and *Pc* (Figure 15.2), part of the negative relationship between these two variables can also be explained mathematically: Increasing *Pc* increases the denominator of *OBA* (number of reported items) without increasing the numerator (number of correct reported items). However, we noted earlier that whereas this expected negative relationship holds for the relatively large

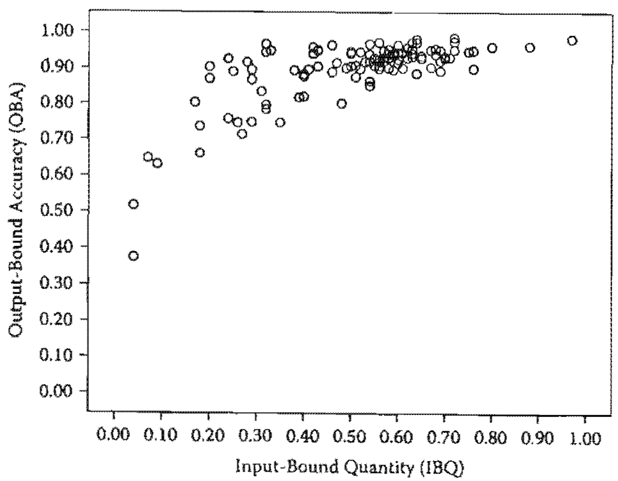


**Figure 15.5** The bivariate distribution of input-bound quantity (*IBQ*) performance and output-bound accuracy (*OBA*) performance scores across the 108 studies.



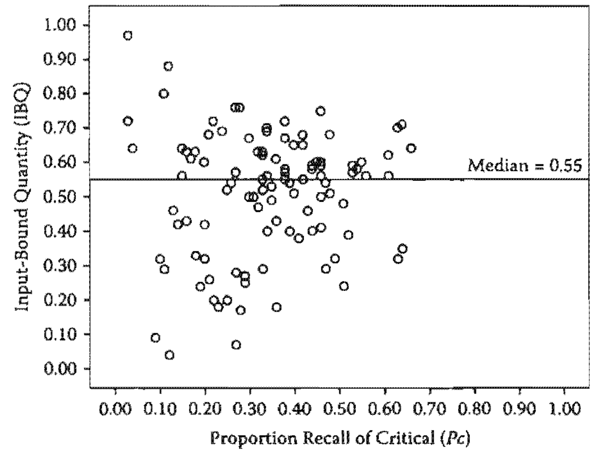


**Figure 15.6** The bivariate distribution of the proportion of recall of the critical nonpresented item (*Pc*) and the input-bound quantity (*IBQ*) performance across the 108 studies. Indicated in the figure is the median of *IBQ* scores.

number of studies that yielded high levels of *OBA* (.90 or above), a surprising positive relationship between *OBA* and *Pc* is observed across the studies yielding lower levels of *OBA*. We speculate that the direction of the *OBA*–*Pc* relationship is in fact moderated by *IBQ*. Figure 15.6 shows that the overall relationship between *Pc* and *IBQ* is also nonlinear ($r = .12$) and inverted U shaped, with a positive relationship observed across the studies exhibiting below-median levels of *IBQ* ($r = .55$), and a negative correlation observed across the studies exhibiting above-median levels of *IBQ* ($r = -.43$). The latter correlation would seem to derive from the positive relationship between memory retention and memory monitoring assumed earlier in explaining the positive *IBQ*–*OBA* relationship. Indeed, in a multiple regression analysis performed across 55 DRM lists, Roediger, Watson, et al. (2001) observed a similar negative correlation ($r = -.43$) between veridical recall of list items (*IBQ*) and false recall of the critical item (*Pc*). They took this correlation to indicate that the better encoded list items are, the more easily they can be distinguished from the illusory critical item. Thus, as *IBQ* (item retention) increases from moderate to high levels, improved monitoring would cause *Pc* to decrease and *OBA* to increase, accounting for the expected negative correlation between them.

For the same reason (dependence of memory monitoring on retention), one would expect *OBA* to increase and *Pc* to decrease in moving from low to moderate levels of *IBQ*. Instead, however, *OBA* and *Pc* jointly increase, yielding the surprising positive correlation between them. We suspect that the anomalous increase in *Pc* stems from the dependence of critical lure production on implicit activations (e.g., Gallo & Roediger, 2002; Roediger, Balota, et al., 2001; Roediger, Watson, et al., 2001) or gist memory (e.g., Brainerd, Wright, Reyna, & Mojardin, 2001): Conditions that yield very low levels of study item memory (*IBQ*) may also yield very low levels of gist or critical lure accessibility (*Pc*), so that the production of both studied items and critical lure items would be jointly impaired. Moreover, if the factors that lead to decreased retention (e.g., longer retention intervals) induce a sharper decline in memory for the studied items than for the critical lures (e.g., McDermott, 1996; Payne et al., 1996; Toglia et al., 1999), a reduced critical lure rate will be accompanied by reduced *OBA*, because the numerator of *OBA* (number of recalled study items) is decreasing at a faster rate than is the denominator (number of reported items, which includes gist-based lures; see the data of Seamon, Luo, Kopecky, et al., 2002, Table 15.2). Finally, although fewer critical lure commission errors were produced under conditions that yielded low *IBQ*, the number of noncritical commission errors actually increased (an average of 2.1 in studies with above-median *IBQ* vs. 3.3 in studies with below-median *IBQ*).[5] This difference, too, would contribute to the dissociation between *OBA* and *Pc*, and again suggests the possible role of gist memory: Because gist memory may support the production of critical lure errors, its decrease over time should reduce the rate of such errors, while at the same time increasing the rate of other, gist-inconsistent commission errors, which otherwise might have been edited out.

Of course, this possible account of the *OBA–Pc* relationship observed across the 108 DRM studies is quite tentative and should be treated primarily as a source of future hypotheses. Regardless of the reason for divergence between *OBA* and *Pc*, the point remains that these measures tell a very different story about the overall reliability of memory in the DRM studies.

## DISCUSSION

Studies using the DRM paradigm have produced extensive evidence that false memories can be readily induced in the laboratory, and that such memories are often endorsed with great confidence. What are the implications of this evidence regarding the trustworthiness of memory reports in general? The high probability of recalling nonpresented words might lead one to place little faith in memory reports. However, our analyses indicate that even under the rather unusual conditions of the DRM paradigm, memory reports as a whole are highly reliable: Each reported item has about a .90 probability of being correct.

In what follows, we comment on three issues: the distinction between input-bound and output-bound perspectives on memory performance, the issue of memory accuracy in the DRM paradigm, and finally, the issue of ecological representativeness of DRM findings.

### *The Distinction Between Input-Bound and Output-Bound Perspectives*

We distinguished three different ways of assessing memory performance in general, and in the DRM paradigm specifically:

1. Input-bound quantity (or accuracy) performance—the probability that any individual studied item will be produced
2. Output-bound accuracy performance—the probability that any individual reported item will be correct
3. Critical error performance—the probability that a particular prespecified commission error will be made

We argued that DRM research has focused on 3, comparing it with 1, essentially ignoring 2. By doing so, the findings are liable to be misinterpreted with respect to 2. The input-bound perspective, reflected in the comparison between 1 and 3, is appropriate given the experimental goals of researchers who use the DRM paradigm to study the mechanisms responsible for this particular type of false memory. For those who are interested in gaining information about the overall dependability of memory reports, however, the output-bound perspective is the one that is directly relevant. Thus, from the vantage point of a judge, police officer, or for that matter, any recipient of information drawn from another person's memory, the crucial question may be: To what extent can one count on each reported item of information to be correct? As we have shown, the same data, examined from these different perspectives, can lead to very different conclusions.

The confusion between the two perspectives is not unlikely to occur, as suggested by research on the inversion of conditional probabilities (e.g., Sherman, McMullen, & Gavanski, 1992). In fact, this confusion may be responsible for our students' faulty conclusions (see Table 15.1) that within the specific conditions of the DRM paradigm, a reported item is as likely, if not more likely, to be wrong than right.

There is no question that the results obtained in the DRM paradigm are striking in showing that under certain conditions, however contrived these may be, participants can be made to recall with a very high probability a particular item that was not presented. As noted earlier, much of the storehouse-guided study of memory has paid little attention to commission errors (Koriat et al., 2000; Roediger, 1996). The DRM paradigm, in contrast, has turned the floodlight precisely on such errors, singling them out as a target of study. Indeed, Roediger and McDermott's (1995) paradigm-setting work, and the vast amount of research and interest that it sparked, has had an immense impact in advancing the study of memory accuracy and error. The message delivered by this research, however, tends to emphasize, explicitly or implicitly, the fragility of memory and the ease with which false memories can be induced. Our analysis indicates that notwithstanding the impressive findings produced by the DRM research, memory is by and large quite accurate even in this paradigm.

Two possible reservations may be raised regarding the high level of output-bound accuracy observed in the DRM paradigm. The first is that this result simply stems from the fact that DRM lists are constructed so that they converge on a single nonpresented item. Therefore, the frequency of producing that single critical commission error is likely to be much lower than the summed frequencies of the many studied items, thereby yielding a high overall *OBA* percentage. The low ratio of commission errors to studied items in the DRM paradigm, however, is not simply a methodological artifact, because it apparently reflects the large amount of "ammunition" (i.e., associated study items) needed to induce a specific commission error. Of course, the ratio of commission errors to studied items might differ both within (as discussed earlier) and between experimental paradigms, and in fact, there are indications that *OBA* is somewhat lower in some free-report paradigms that are used to induce memory errors (e.g., Kelley & Sahakyan, 2003; Pansky & Koriat, 2004; Sommers & Lewis, 1999; Toglia et al., 1999). Nevertheless, the fact remains that *OBA* is actually very high in a paradigm whose implicit message is that memory reports are not to be trusted.

A second reservation is that in our analyses we adopted the assumption underlying the input-bound, quantity-oriented approach to memory: that all items are interchangeable, that is, equivalent, as far as memory performance is concerned (see Koriat & Goldsmith, 1996a). Indeed, the assessment procedures we used, like those characteristic of much of memory research, embody the assumption that what matters is not *what* is remembered or misremembered, but rather *how much*. However, one may envisage situations in which certain commission errors are especially critical even if their contribution to the overall output-bound accuracy of the report is negligible. For example, many elements of a crime episode might converge in suggesting the presence of a particular weapon, even though no such weapon was there. A single commission error concerning the presence of the weapon could have tragic consequences, even if a large amount of correct information was also remembered and reported. Clearly, as we readily acknowledge, in applied settings there is more to the assessment of false recall than is captured by the output-bound accuracy percentage alone (for related discussion, see Fisher, 1996; Goldsmith & Koriat, 1996).



### *Output-Bound Memory Accuracy and Its Underlying Mechanisms*

The positive message delivered by the present analysis is that when people attempt to provide an accurate account of what has occurred, their freely volunteered memory reports are by and large correct. One reason for this is that output-bound memory accuracy performance is—to a large extent—under strategic control: Regardless of how much information one "remembers," one can still boost one's accuracy to relatively high levels by volunteering only information that one is sure about and screening out information that is likely to be wrong. As noted earlier, according to Koriat and Goldsmith's (1996b) model, metacognitive monitoring and control processes play a crucial role in this regulation. Hence, the level of accuracy that is attained depends heavily on the effectiveness of these processes.

In terms of that model, the occurrence of false recalls in the DRM paradigm reflects not only a memory failure but also a failure of metamemory (Roediger & Gallo, 2004). Although the structure of the DRM list increases the likelihood that the critical lure will come to mind as a response candidate during recall, in principle, effective monitoring and control processes could operate to reject that candidate once it comes to mind. For example, Brainerd, Reyna, Wright, and Mojardin (2003) postulate an editing process called "recollection rejection," in which distracters that are consistent with the gist of a presented item are rejected when the verbatim trace of that item is accessed. However, as noted by Gallo (2004), critical DRM errors cannot usually be identified by such a process, because accessing the verbatim traces of some or even most of the studied items does not exclude (disqualify) the possible co-occurrence of the critical lure item.

Apparently, then, the DRM paradigm creates a situation in which monitoring and control processes are relatively ineffective in editing out the critical lure. Indeed, warning participants about the DRM effect and instructing them to avoid reporting nonstudied but related words yields only negligible reductions in false recall and recognition of the critical lures (Gallo, Roediger, & McDermott, 2001b; McDermott & Roediger, 1998; Neuschatz, Payne, Lampinen, & Toglia, 2001). Moreover, recollections of the critical lures are often experienced as phenomenologically compelling (e.g., Norman & Schacter, 1997; Roediger & McDermott, 1995), and are therefore volunteered under free-report conditions. Koriat and Goldsmith's (1996b) results suggest that rememberers' decision to volunteer or withhold a candidate answer in free recall depends almost entirely on subjective confidence in its correctness, with within-participant gamma correlations between confidence and volunteering averaging over .95 in many experiments (see also Kelley & Sahakyan, 2003). Thus, the observation that participants often endorse the critical lure with high confidence (Roediger & McDermott, 1995; Payne et al., 1996) implies that these errors are unlikely to be selectively omitted from rememberers' memory reports.

In regulating the accuracy of what they report, however, rememberers have more options available to them than what has been discussed so far. Another means of strategic regulation that is perhaps more generally available in real-life memory situations is control over the precision or *grain size* of the information that is reported. For example, rememberers may report "in the late afternoon" rather than "at 4:00," or a "fruit" rather than an "apple" (see Goldsmith et al., 2005; Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Weber & Brewer, 2008). Neisser (1988) observed that when answering open-ended questions, participants tend to provide answers at a level of generality at which they are not likely to be mistaken. Goldsmith et al. (2002, 2005) found that when participants are allowed to control the grain size of their report, they do so in a strategic manner, sacrificing informativeness (degree of precision) for the sake of accuracy when their subjective confidence in the more precise, informative answer is low (but for a somewhat more complex view, see Ackerman & Goldsmith, 2008).

Control over grain size is denied in the DRM paradigm, as well as in almost all list-learning memory tasks. If such control were allowed in DRM studies, however, *OBA* would undoubtedly be even higher than what was observed in our earlier analysis. The irony with regard to the DRM paradigm is particularly poignant: Gist is used in the DRM paradigm to create memory errors, whereas in most real-life situations rememberers use gist to avoid them. Indeed, some, if not most, of the commission errors in memory studies represent partial recalls, as when a person recalls *pants* instead of *jeans* (Pansky & Koriat, 2004) or when some information is retained but its source is lost (see Mitchell & Johnson, 2000, for a review). Clearly, the definition of what constitutes false memory is not simple, as reflected in the criteria that must sometimes be set by the experimenter in testing and scoring memory performance (e.g., Ebbesen & Rienick, 1998; Koriat, Levy-Sadot, Edry, & de Marcas, 2003).



In sum, the DRM paradigm appears to yield a troublesome combination of memory and monitoring impairment. However, this combination occurs specifically for the critical lure, as suggested by the observation that the rate of producing this particular commission error was higher than the overall rate of all *other* commission errors. Therefore, the output-bound accuracy for the list as a whole remains quite high. Furthermore, even the critical lure errors produced in this paradigm are in some sense "correct"—at a higher, gist level. This raises further questions regarding memory accuracy, and how it should be assessed (see also Spence, 1982).

### *The Issue of Ecological Representativeness*

At the opening of this chapter, we stated that we would put the "ecological representativeness" issue aside and examine performance in the DRM paradigm as if it is in fact representative of real-life memory situations. In this final section, we briefly address the issue of ecological representativeness.

Some of the recent work on false memory illustrates a distinction between two sometimes conflicting objectives of cognitive research (cf. Chomsky, 1965). One objective is to *describe* the state of affairs in the real world. For example, researchers may wish to delineate the strengths and weaknesses of cognitive abilities, to specify how memory performance changes as a function of retention interval, to evaluate the veracity of human memory under different conditions, and to describe the various biases that affect performance. Researchers with this agenda, by definition, should restrict themselves to conditions that are ecologically representative. The plea for representative research design has been voiced most strongly by Brunswik (1956; see also Gibson, 1979; Gigerenzer, Hoffrage, & Kleinbölting, 1991). The question of whether data should be collected only in the laboratory or also in naturalistic settings has been a subject of some dispute (e.g., Banaji & Crowder, 1989; Neisser, 1988; see special issue of *American Psychologist*, 1991), but it is clear that to be descriptive of the actual magnitudes of variables and their relationships in the real world, the experimental conditions must be representative of conditions and variations in the real world.

The second objective is theoretically oriented: It is aimed at *explaining* the phenomena under investigation and gaining an *understanding* of their underlying mechanisms. This objective is illustrated by research that attempts to clarify the processes that cause forgetting, or those that underlie false memories. Research carried out within this agenda need not respect the plea for representative design and need not confine itself to the conditions that approximate real-world settings. In fact, it is sometimes precisely under extreme or deviant conditions that one is best able to gain an understanding of the normal processes that occur under more natural conditions. As noted by Roediger (1996), this approach has been quite effective in the area of perception, in which the study of perceptual illusions induced under unusual conditions such as the trapezoidal window, or the distorted room (see Ittelson, 1952), has revealed important principles about perception in general that are not transparent when ordinary perceptual processes are examined directly. Similarly, the many studies of memory and metamemory illusions (see Koriat, 2007, 2008) provide valuable information precisely because they succeed in decoupling processes and effects that generally go hand in hand under normal conditions. Although much of that research has focused on phenomena testifying to the existence of illusions and errors, other research has investigated cases of exceptionally good memory performance (e.g. Ericsson, Charness, Feltovich, & Hoffman, 2006), which has also contributed to our understanding of basic cognitive processes. Clearly, the significant contribution of the extensive research with the DRM paradigm lies within the explanatory agenda—providing important insights about the processes that lead to false memories and the subjective qualities of veridical and false memories.

However, while the focus on ecologically deviant conditions can have important theoretical benefits, it also holds the danger that the research results might be misinterpreted as having descriptive relevance. This danger is twofold. First, the frequent sampling of conditions that yield illusions and errors may create the unintended impression that the frequency of the phenomena in the experimental research literature mirrors their frequency in the real world. This impression may be amplified by the salience of studies showing surprisingly high levels of false memory. Thus, for example, the availability heuristic (Tversky & Kahneman, 1973) could lead memory researchers, as well as the general public, to form a biased judgment regarding memory performance that overemphasizes the sins of memory—the preponderance of error, illusions, and false memory. Second, the focus on distortion and error, and the ensuing challenges in clarifying the mechanisms that may engender or prevent such frailties, may lead to a preoccupation with the explanatory objective at the expense of the descriptive objective. As the results of the present study suggest, theoretically oriented researchers who do not also attend to the descriptive message of their research may unintentionally convey an incomplete or distorted impression of memory performance in real-world contexts.



Having said that, we should emphasize that the mere fact that people can be made to misremember, and to do so with high confidence, is a message of great practical importance with regard to the way in which witness testimony should be treated, as well as in other memory contexts. However, it would seem that the time has come to try to refine that message for the sake of legal practitioners and others. What is the likelihood that any given witness statement is true or false? The output-bound accuracy measure applied to results obtained in both explanatory- and descriptive-oriented research, including the DRM paradigm, suggests that under conditions of free reporting, the dependability of reported information is quite high. Nevertheless, much work remains to be done in identifying the conditions that increase false reports, so that these can be taken into account in attempting to ascertain the reliability of memory in specific reporting contexts.

## ACKNOWLEDGMENTS

This work was supported by a grant from the European Commission (FP6-NEST: EYEWITMEM; 43460). We thank Einat Tenenboim for her valuable help in collecting the data.

Address correspondence to Asher Koriat, Department of Psychology, University of Haifa, Haifa 31905, Israel. E-mail: akoriat@research.haifa.ac.il.

## ENDNOTES

1. Because of the traditional concern with memory quantity, the common practice in memory research has been to ignore commission errors altogether (see Roediger, 1996). As a consequence, it is not possible to determine the output-bound accuracy observed in many of the reported studies. However, with the growing interest in memory accuracy, an increasing number of studies either report output-bound measures directly or provide the data from which these measures can be calculated. The latter generally applies to studies reporting recall results in the context of the DRM paradigm.
2. The 108 studies include all those that we were able to identify until 2007. *Pc* was not reported for three of these studies (studies 9 to 11 in Table 15.2). Therefore, the analyses involving *Pc* are based on 105 studies.
3. Note that for our current purposes inferential statistics are not used or needed, because we are limiting ourselves to a comparison of trends involving *OBA*, *IBQ*, and *Pc* in the current (very large) sample of DRM studies. In any case, the use of meta-analytic inferential statistics was precluded by the preponderant lack of information regarding the variance in *OBA* that was observed within the individual studies.

4. Note that the high variability in the *OBA* scores of amnesic patients can be partially attributed to the pooling together of different types of amnesia (e.g., amnesic patients with frontal lobe damage vs. patients with damage in the medial temporal lobe or diencephalic region), each of which exhibits a different pattern of performance in the DRM paradigm (see Melo, Winocur, & Moscovitch, 1999; Schacter et al., 1996).
5. Note that this difference also holds when *Nnc* is calculated as a percentage of the number of studied items (mean = 1.6% in studies with above-median *IBQ* vs. 2.7% in studies with below-median *IBQ*).

## REFERENCES


Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting—With and without satisfying knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1224–1245.

*American Psychologist.* (1991). *46*(1).

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist, 44,* 1185–1193.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Basden, B. H., Basden, D. R., Thomas, R. L., & Souphasith, S. (1998). Memory distortion in group recall. *Current Psychology: Developmental, Learning, Personality, Social, 16,* 225–246.

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 941–947.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110,* 762–784.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 307–327.

Brunswik, E. (Ed.). (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Dehon, H., & Bredart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology and Aging, 19,* 191–197.

Dewhurst, S. A., & Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychological Science, 15,* 782–786.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83,* 745–762.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.). (2006). *The Cambridge handbook of expertise and expert performance.* New York: Cambridge University Press.

Fisher, R. P. (1996). Misconceptions in design and analysis of research with the cognitive interview. *Psycholoquy, 7.* Retrieved from witness-memory.12.fisher

Freyd, J. J., & Gleaves, D. H. (1996). "Remembering" words not presented in lists: Relevance to the current recovered/false memory controversy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 811–813.

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 120–128.

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 339–353.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47,* 469–497.

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin and Review, 8,* 579–586.

Geraerts, E., Smeets, E., Jelicic, M., van Heerden, J., & Merckelbach, H. (2005). Fantasy proneness, but not self-reported trauma is related to DRM performance of women reporting recovered memories of childhood sexual abuse. *Consciousness and Cognition: An International Journal, 14,* 602–612.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin and Company.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Goldsmith, M., & Koriat, A. (1996). The assessment and control of memory accuracy: Commentary on Memon & Stevenage on witness-memory. *Psycholoquy, 7.* Retrieved from witness-memory.9.goldsmith

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin & B. Ross (Eds.), *Psychology of learning and motivation: Memory use as skilled cognition* (Vol. 48, pp. 1–60). San Diego, CA: Elsevier.

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52,* 505–525.

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131,* 73–95.

Harbluk, J. L., & Weingartner, H. J. (1997). Memory distortions in detoxified alcoholics. *Brain and Cognition, 35,* 328–330.

Higham, P. A. (2007). No Special K! A signal-detection framework of the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General, 136*, 1–22.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1999). Age, testing at preferred or nonpreferred times (testing optimality), and false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 23–40.

Ittelson, W. H. (Ed.). (1952). *The Ames demonstrations in perception; a guide to their construction and use*. Princeton, NJ: Princeton University Press.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48*, 704–721.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). New York: Cambridge University Press.

Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 945–959.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297–315.

Koriat, A., & Goldsmith, M. (1996a). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences, 19*, 167–228.

Koriat, A., & Goldsmith, M. (1996b). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–537.

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. (2001). The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology, 79*, 405–437.

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1095–1105.

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin and Review, 6*, 130–135.

Libby, L. K., & Neisser, U. (2001). Structure and strategy in the associative false memory paradigm. *Memory, 9*, 145–163.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.



Loftus, E. F. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews Neuroscience, 4*, 231–234.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*, 212–230.

McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39*, 508–520.

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 45*, 160–176.

McKelvie, S. (2001). Effects of free and forced retrieval instructions on false recall and recognition. *Journal of General Psychology, 128*, 261–278.

Melo, B., Winocur, G., & Moscovitch, M. (1999). False recall and false recognition: An examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. *Cognitive Neuropsychology, 16*, 343–359.

Milani, R., & Curran, H. V. (2000). Effects of a low dose of alcohol on recollective experience of illusory memory. *Psychopharmacology, 147*, 397–402.

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review, 106*, 398–405.

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). Oxford, UK: Oxford University Press.

Neisser, U. (1988). Time present and time past. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 2, pp. 545–560). Chichester, UK: Wiley.

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9*, 53–71.

Newstead, B. A., & Newstead, S. E. (1998). False recall and false memory: The effects of instructions on memory errors. *Applied Cognitive Psychology, 12*, 67–79.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory and Cognition, 25*, 838–848.

Pansky, A., & Koriat, A. (2004). The basic-level convergence effect in memory distortions. *Psychological Science, 15*, 52–59.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness recall and testimony. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 93–150). New York: Guilford Press.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*, 261–285.

Poole, D. A., & White, L. T. (1993). Two years later: Effect of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology, 29*, 844–853.

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin and Review, 3*, 105–111.

Rhodes, M. G., & Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin and Review, 7*, 158–162.

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*, 231–237.

Roebers, C. M., & Fernandez, O. (2002). The effects of accuracy motivation on children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development, 3*, 415–443.

Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory and Cognition, 8*, 231–246.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language, 35*, 76–100.

Roediger, H. L., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association.

Roediger, H. L., & Gallo, D. A. (2004). Associative memory illusions. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 309–326). Hove, UK: Psychology Press.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Roediger, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9*, 123–127.

Roediger, H. L., McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory* (Vol. 2, pp. 187–245). Hove, UK: Psychology Press.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin and Review, 8*, 385–407.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton Mifflin Co.

Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language, 35*, 319–334.



Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory and Cognition, 30*, 1054–1064.

Seamon, J. G., Luo, C. R., Shulman, E. P., Toner, S. K., & Caglar, S. (2002). False memories are hard to inhibit: Differential effects of directed forgetting on accurate and false recall in the DRM procedure. *Memory, 10*, 225–238.

Sherman, S. J., McMullen, M. N., & Gavanski, I. (1992). Natural sample spaces and the inversion of conditional judgments. *Journal of Experimental Social Psychology, 28*, 401–421.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin and Review, 5*, 710–715.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, 40*, 83–108.

Spence, D. P. (1982). *Narrative truth and historical truth*. New York: Norton.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*, 233–256.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*, 230–241.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232.

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied, 14*, 50–60.

Winograd, E., Peluso, J. P., & Glover, T. A. (1998). Individual differences in susceptibility to memory illusions. *Applied Cognitive Psychology, 12*, S5–S27.

Zoellner, L. A., Foa, E. B., Brigidi, B. D., & Przeworski, A. (2000). Are trauma victims susceptible to "false memories"? *Journal of Abnormal Psychology, 109*, 517–524.

Psychology Press
Taylor & Francis Group
270 Madison Avenue
New York, NY 10016

Psychology Press
Taylor & Francis Group
27 Church Road
Hove, East Sussex BN3 2FA



Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-84872-891-2 (Hardback)



**Library of Congress Cataloging-in-Publication Data**

Successful remembering and successful forgetting : a festschrift in honor of Robert A. Bjork / edited by Aaron S. Benjamin.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84872-891-2 (hardcover : alk. paper)
1. Memory. I. Bjork, Robert A. II. Benjamin, Aaron S.

BF371.S86 2011
153.1'2--dc22 2010041648

**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**and the Psychology Press Web site at**
**http://www.psypress.com**

# Successful Remembering and Successful Forgetting

## A Festschrift in Honor of Robert A. Bjork

EDITED BY **Aaron S. Benjamin**

Psychology Press
Taylor & Francis Group
New York London

**Table 15.2** Summary Data for 108 DRM Studies

| | Source | Specification | Ps (IBQ) | Pc | OBA |
|---|---|---|---|---|---|
| 1 | Basden et al. (1998) | Experiment 1, nominal condition[v,j] | .46 | .43 | .89 |
| 2 | Basden et al. (1998) | Experiment 1, collaborative condition[v,j] | .40 | .44 | .82 |
| 3 | Dehon & Bredart (2004) | Experiment 1, young adults | .69 | .24 | .95 |
| 4 | Dehon & Bredart (2004) | Experiment 1, older adults | .51 | .48 | .88 |
| 5 | Dehon & Bredart (2004) | Experiment 2, young adults, unwarned | .63 | .16 | .95 |
| 6 | Dehon & Bredart (2004) | Experiment 2, young adults, warned | .64 | .04 | .98 |
| 7 | Dehon & Bredart (2004) | Experiment 2, older adults, unwarned | .40 | .34 | .88 |
| 8 | Dehon & Bredart (2004) | Experiment 2, older adults, warned | .40 | .39 | .88 |
| 9 | Dewhurst & Robinson (2004) | 5 year olds | .31 | . | .84 |
| 10 | Dewhurst & Robinson (2004) | 8 year olds | .54 | . | .85 |
| 11 | Dewhurst & Robinson (2004) | 11 year olds | .69 | . | .92 |
| 12 | Gallo, McDermott et al. (2001) | Experiment 1, auditory presentation | .60 | .46 | .94 |
| 13 | Gallo, McDermott et al.(2001) | Experiment 1, visual presentation[v] | .58 | .38 | .95 |
| 14 | Geraerts et al. (2005) | Recovered memory group, neutral words[v] | .56 | .61 | .92 |
| 15 | Geraerts et al. (2005) | Recovered memory group, trauma-related words[v] | .42 | .20 | .94 |
| 16 | Geraerts et al. (2005) | Repressed memory group, neutral words[v] | .56 | .46 | .93 |
| 17 | Geraerts et al. (2005) | Repressed memory group, trauma-related words[v] | .43 | .16 | .95 |
| 18 | Geraerts et al. (2005) | Continuous memory group, neutral words[v] | .55 | .42 | .93 |
| 19 | Geraerts et al. (2005) | Continuous memory group, trauma-related words[v] | .42 | .14 | .96 |
| 20 | Geraerts et al. (2005) | Control group, neutral words[v] | .59 | .44 | .94 |
| 21 | Geraerts et al. (2005) | Control group, trauma-related words[v] | .46 | .13 | .96 |
| 22 | Harbluk & Weingartner (1997) | Detoxified alcoholics | .59 | .46 | .90 |
| 23 | Harbluk & Weingartner (1997) | Control condition | .68 | .48 | .94 |
| 24 | Intons-Peterson et al. (1999) | Experiment 1, older adults | .39 | .52 | .82 |
| 25 | Intons-Peterson et al. (1999) | Experiment 1, young adults | .56 | .56 | .90 |
| 26 | Intons-Peterson et al. (1999) | Experiment 2, older adults | .35 | .64 | .75 |
| 27 | Intons-Peterson et al. (1999) | Experiment 2, young adults | .60 | .55 | .91 |
| 28 | Intons-Peterson et al. (1999) | Experiment 3, older adults, pictorial presentation[v,j] | .55 | .38 | .91 |
| 29 | Intons-Peterson et al. (1999) | Experiment 3, young adults, pictorial presentation[v,j] | .61 | .17 | .94 |
| 30 | Lampinen et al. (1999) | Experiment 1 | .61 | .36 | .90 |
| 31 | Lampinen et al. (1999) | Experiment 2 | .56 | .34 | .91 |
| 32 | Libby & Neisser (2001) | Experiment 1, distraction at study, short list | .76 | .28 | .90 |
| 33 | Libby & Neisser (2001) | Experiment 1, distraction at study, long list | .55 | .33 | .93 |
| 34 | Libby & Neisser (2001) | Experiment 1, rehearsal at study, short list | .88 | .12 | .96 |
| 35 | Libby & Neisser (2001) | Experiment 1, rehearsal at study, long list | .57 | .38 | .92 |
| 36 | McDermott (1996) | Experiment 1, immediate testing | .58 | .44 | .93 |
| 37 | McDermott (1996) | Experiment 1, 30-second delayed testing | .50 | .46 | .91 |
| 38 | McDermott (1996) | Experiment 1, 48-hour delayed testing[j] | .04 | .12 | .52 |
| 39 | McKelvie (2001) | Experiment 1 | .70 | .34 | .93 |
| 40 | Melo et al. (1999) | MTL/D amnesic patients | .32 | .63 | .80 |
| 41 | Melo et al. (1999) | FL amnesic patients | .24 | .19 | .93 |
| 42 | Melo et al. (1999) | FL nonamnesic patients | .41 | .46 | .90 |
| 43 | Melo et al. (1999) | Control participants | .49 | .35 | .90 |
| 44 | Milani & Curran (2000) | Alcohol consumption condition | .51 | .40 | .91 |
| 45 | Milani & Curran (2000) | Placebo consumption condition | .54 | .39 | .92 |
| 46 | Miller & Wolford (1999) | Experiment 1 | .76 | .27 | .95 |
| 47 | Miller & Wolford (1999) | Experiment 2 | .68 | .42 | .96 |

*—continued*

**Table 15.2 (continued)** Summary Data for 108 DRM Studies

| | | | | | |
|---|---|---|---|---|---|
| 48 | Neuschatz et al. (2001) | Experiment 2[j] | .32 | .49 | .79 |
| 49 | Newstead & Newstead (1998) | Pilot experiment, 15 year olds | .75 | .46 | .95 |
| 50 | Newstead & Newstead (1998) | Main experiment, 13–16 year olds | .72 | .38 | .95 |
| 51 | Norman & Schacter (1997) | Experiment 1, older adults | .48 | .51 | .80 |
| 52 | Norman & Schacter (1997) | Experiment 1, young adults | .67 | .38 | .90 |
| 53 | Norman & Schacter (1997) | Experiment 2, older adults | .54 | .47 | .86 |
| 54 | Norman & Schacter (1997) | Experiment 2, young adults | .69 | .34 | .89 |
| 55 | Payne et al. (1996) | Experiment 1 | .60 | .45 | .93 |
| 56 | Read (1996) | Experiment 1 | .64 | .66 | .89 |
| 57 | Rhodes & Anastasi (2000) | Experiment 1, deep encoding[j] | .29 | .47 | .75 |
| 58 | Rhodes & Anastasi (2000) | Experiment 1, shallow encoding[j] | .18 | .23 | .74 |
| 59 | Rhodes & Anastasi (2000) | Experiment 2, deep encoding[j] | .38 | .41 | .89 |
| 60 | Rhodes & Anastasi (2000) | Experiment 2, shallow encoding[j] | .09 | .09 | .63 |
| 61 | Robinson & Roediger (1997) | Experiment 1, 3 associates per list[v] | .97 | .03 | .98 |
| 62 | Robinson & Roediger (1997) | Experiment 1, 6 associates per list[v] | .80 | .11 | .96 |
| 63 | Robinson & Roediger (1997) | Experiment 1, 9 associates per list[v] | .68 | .21 | .95 |
| 64 | Robinson & Roediger (1997) | Experiment 1, 12 associates per list[v] | .57 | .27 | .95 |
| 65 | Robinson & Roediger (1997) | Experiment 1, 15 associates per list[v] | .50 | .31 | .95 |
| 66 | Robinson & Roediger (1997) | Experiment 2, 3 associates and 12 fillers per list[v] | .72 | .03 | .99 |
| 67 | Robinson & Roediger (1997) | Experiment 2, 6 associates and 9 fillers per list[v] | .64 | .15 | .97 |
| 68 | Robinson & Roediger (1997) | Experiment 2, 9 associates and 6 fillers per list[v] | .60 | .20 | .96 |
| 69 | Robinson & Roediger (1997) | Experiment 2, 12 associates and 3 fillers per list[v] | .52 | .25 | .95 |
| 70 | Robinson & Roediger (1997) | Experiment 2, 15 associates per list[v] | .50 | .30 | .94 |
| 71 | Roediger & McDermott (1995) | Experiment 1 | .65 | .40 | .94 |
| 72 | Schacter et al. (1996) | Amnesic patients | .27 | .29 | .72 |
| 73 | Schacter et al. (1996) | Control participants | .52 | .33 | .90 |
| 74 | Seamon, Luo, Kopecky et al. (2002) | Experiment 1, immediate testing[j] | .17 | .28 | .80 |
| 75 | Seamon, Luo, Kopecky et al. (2002) | Experiment 1, 2-week delayed testing[j] | .07 | .27 | .65 |
| 76 | Seamon, Luo, Kopecky et al. (2002) | Experiment 1, 2-month delayed testing[j] | .04 | .12 | .38 |
| 77 | Seamon, Luo, Shulman et al. (2002) | Remember/remember instructions, 8 lists[v,j] | .28 | .27 | .92 |
| 78 | Seamon, Luo, Shulman et al. (2002) | Forget/remember instructions, 8 lists[v,j] | .25 | .29 | .89 |
| 79 | Seamon, Luo, Shulman et al. (2002) | Remember/remember instructions, 12 lists[v,j] | .20 | .22 | .90 |
| 80 | Seamon, Luo, Shulman et al. (2002) | Forget/remember instructions, 12 lists[v,j] | .20 | .25 | .87 |
| 81 | Seamon et al. (2003) | Experiment 1, only hear the associates condition | .67 | .30 | .95 |
| 82 | Seamon et al. (2003) | Experiment 1, write the associates condition | .63 | .18 | .97 |
| 83 | Seamon et al. (2003) | Experiment 1, write the second letter of the associates condition | .56 | .15 | .97 |
| 84 | Seamon et al. (2003) | Experiment 1, count back by threes condition | .43 | .36 | .90 |
| 85 | Smith & Hunt (1998) | Experiment 1, auditory presentation[j] | .26 | .21 | .75 |
| 86 | Smith & Hunt (1998) | Experiment 1, visual presentation[v,j] | .29 | .11 | .87 |
| 87 | Smith & Hunt (1998) | Experiment 2, auditory presentation | .65 | .42 | .93 |
| 88 | Smith & Hunt (1998) | Experiment 2, visual presentation[v] | .72 | .22 | .97 |
| 89 | Smith & Hunt (1998) | Experiment 3, auditory presentation, pleasantness ratings[j] | .32 | .20 | .94 |
| 90 | Smith & Hunt (1998) | Experiment 3, auditory presentation, standard encoding[j] | .29 | .33 | .90 |
| 91 | Smith & Hunt (1998) | Experiment 3, visual presentation, pleasantness ratings[v,j] | .32 | .10 | .97 |
| 92 | Smith & Hunt (1998) | Experiment 3, visual presentation, standard encoding[v,j] | .33 | .18 | .95 |
| 93 | Sommers & Lewis (1999) | Experiment 1, phonological associates | .58 | .54 | .94 |
| 94 | Sommers & Lewis (1999) | Experiment 2, phonological associates, single speaker | .62 | .61 | .93 |

—continued

**Table 15.2 (continued)** Summary Data for 108 DRM Studies

| | | | | | |
|---|---|---|---|---|---|
| 95 | Sommers & Lewis (1999) | Experiment 2, phonological associates, multiple speakers (blocked) | .71 | .64 | .93 |
| 96 | Sommers & Lewis (1999) | Experiment 2, phonological associates, multiple speakers (random) | .70 | .63 | .93 |
| 97 | Sommers & Lewis (1999) | Experiment 3, phonologically most confusable associates | .57 | .53 | .93 |
| 98 | Sommers & Lewis (1999) | Experiment 3, phonologically least confusable associates | .62 | .33 | .96 |
| 99 | Toglia et al. (1999) | Experiment 2, blocked lists[j] | .24 | .51 | .76 |
| 100 | Toglia et al. (1999) | Experiment 2, randomly mixed lists[j] | .18 | .36 | .66 |
| 101 | Tun et al. (1998) | Experiment 1, older adults | .53 | .35 | .92 |
| 102 | Tun et al. (1998) | Experiment 1, young adults | .63 | .33 | .93 |
| 103 | Tun et al. (1998) | Experiment 2, older adults | .47 | .32 | .92 |
| 104 | Tun et al. (1998) | Experiment 2, young adults | .63 | .32 | .94 |
| 105 | Winograd et al. (1998) | | .58 | .44 | .90 |
| 106 | Zoellner et al. (2000) | Traumatized PTSD participants | .54 | .47 | .94 |
| 107 | Zoellner et al. (2000) | Traumatized non-PTSD participants | .59 | .53 | .94 |
| 108 | Zoellner et al. (2000) | Control participants | .54 | .26 | .97 |

[v] Visual presentation at study; otherwise, auditory presentation at study.

[j] Joint recall test following the presentation of all the lists; otherwise, recall test following each list.