Journal of Experimental Psychology:
Learning, Memory, and Cognition
1998, Vol. 24, No. 1, 3–26

0278-7393/98/$3.00

Invariance in Automatic Influences of Memory: Toward a User's Guide for the Process-Dissociation Procedure

Larry L. Jacoby
McMaster University

Three experiments investigated assumptions of the process-dissociation procedure for separating consciously controlled and automatic influences of memory. Conditions that encouraged direct retrieval revealed process dissociations. Manipulating attention during study or manipulating study time affected recollection but left automatic influences of memory relatively invariant. However, paradoxical dissociations were found when conditions encouraged use of a generate–recognize strategy, violating assumptions underlying the estimation procedure. Use of subjective reports to gain estimates produced parallel results. Easily observed correlations are shown to be not useful for testing assumptions underlying the process-dissocation procedure. A multinomial model produced results that agree with those from the process-dissociation approach.

The process-dissociation procedure (Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993) was introduced as a general method for separating the contributions of consciously controlled processes, such as recollection, from those of unconscious or automatic uses of memory. The procedure builds on findings of dissociations between performance on direct and indirect tests of memory (for a review, see Roediger & McDermott, 1993). However, rather than identifying recollection and automatic influences of memory with performance on direct and indirect tests, respectively, the procedure is designed to separate the within-task contributions of the two bases for responding. For example, Jacoby et al. (1993) used the process-dissociation procedure to show that recall cued with word stems, a direct test of memory, involves not only recollection but also reflects automatic influences of memory or implicit memory of the sort measured by using stem-completion performance as an indirect test of memory.

There has been a great deal of controversy surrounding the assumptions underlying the process-dissociation procedure (e.g., Curran & Hintzman, 1995; Graf & Komatsu, 1994; Joordens & Merikle, 1993). The evolution of a theory or methodology is to unveil its assumptions and to come up with ways to test them directly. In the absence of direct tests, evaluation of the new approach must lie in its success in leading to experiments that uncover new empirical facts along with replicable empirical relations. An equally important component in the evolution of a theory or methodology is the discovery of boundary conditions for its applicability. The existence of boundary conditions does not mean that the approach should be discarded entirely. That is, one should not advocate abandoning analysis of variance (ANOVA) just because the variances across conditions are sometimes not homogeneous or the scores are sometimes not normally distributed. Indeed, advances occur when such boundary conditions are discovered. Establishment of boundary conditions allows one to avoid inappropriate application of the procedure.

Of course, as a new approach receives extensive criticism, there is a tendency to simply abandon the approach, although that is not what critics of the approach intended. Against abandoning the process-dissociation approach, the experiments reported here replicated findings offered as support for the approach and identified boundary conditions for those findings. Yet, problems for the process-dissociation approach remain. Those remaining problems are discussed in the broader context of other approaches that share the goal of separating the contributions of different forms or uses of memory to performance of a task.

The Process-Dissociation Procedure: Underlying Assumptions

An experiment done by Jacoby et al. (1993, Experiment 1b) illustrated the process-dissociation procedure. They examined effects of full versus divided attention during study on recall cued with word stems (e.g., *mot*__ for *motel*). For an *inclusion test,* participants were instructed to use the stem as a cue to recall an old word or, if they could not do so, to complete the stem with the first word that came to mind. For an *exclusion test,* participants were instructed to use the stem as a cue to recall an old word but not to use recalled

---

Larry L. Jacoby, Department of Psychology, McMaster University, Hamilton, Ontario, Canada.

This research was supported by Grant OGP0000281 from the Natural Sciences and Engineering Research Council of Canada and Grant SBR-9596209 from the National Science Foundation.

The author expresses appreciation to Gay Snodgrass for discussions about multinomial models and to Colleen Kelley for her comments on earlier versions of this article. Also, thanks to Ann Hollingshead and Jim Debner for assistance during all phases of the research reported here.

Correspondence concerning this article should be addressed to Larry L. Jacoby, who is currently at the Department of Psychology, New York University, 6 Washington Place, New York, New York 10003-6634. Electronic mail may be sent via Internet to larry@xp.psych.nyu.edu.

words to complete the stems. That is, participants were instructed to exclude old words and to complete stems only with unrecalled words. For both inclusion and exclusion tests, participants were correctly informed that many of the stems could be completed only with new words and, so, should be completed with the first word that came to mind. Completion rates for stems corresponding to new words served as an index of base rate against which automatic influences of memory resulting from study were measured.

An inclusion test is like a standard test of cued recall with instructions to guess when unable to recollect. People could complete a stem with an old word either because they recollected the studied word, with a probability of R, or because the old word came automatically to mind, with a probability of A. If these two bases of responding are independent, then inclusion performance equals $R + A - RA$. For the exclusion test, in contrast, participants would complete a stem with an old word only if the word came automatically to mind without recollection of its prior presentation: $A(1 - R) = A - RA$. The difference between the inclusion and exclusion tests provides an estimate of the probability of recollection: R = Inclusion − Exclusion. Given that estimate, one can compute the probability of an old word automatically coming to mind: A = Exclusion/$(1 - R)$.

Using these equations, Jacoby et al. (1993) showed that dividing attention during study reduced R (.25 vs. .00) but left A almost invariant (.47 vs. .46). That is, the estimates showed a process dissociation similar to the task dissociations found between direct and indirect memory tests (Koriat & Feuerstein, 1976; Parkin, Reid, & Russo, 1990). That dissociation provides support for the independence assumption underlying the process-dissociation procedure by showing that a manipulation traditionally identified with cognitive control selectively affects the estimate of consciously controlled processes. Similar process dissociations have been found in several other experiments (for a review, see Jacoby, Yonelinas, & Jennings, 1997).

In addition to the independence assumption, the estimation procedure rests on the assumption that R is equal for the inclusion and exclusion tests. It is also assumed that A is equal for the two types of test. To assess automatic influences of memory, Jacoby et al. (1993) compared A with the base rate of completing stems corresponding to new words (.46 vs. .35) and found a significant difference. It was important that base rates did not differ significantly across the inclusion and exclusion tests or across the manipulation of full versus divided attention. As described later, differences in base rates can reflect the violation of assumptions underlying the process-dissociation procedure.

Generate–Recognize Versus Direct Retrieval: Boundary Conditions for Independence

Details of test instructions potentially serve as an important boundary condition for finding process dissociations. The instructions used by Jacoby et al. (1993) were meant to encourage participants to retrieve directly earlier-studied words, using word stems as cues. To satisfy assumptions underlying the process-dissociation procedure, they required participants to exclude old words only on the basis of recollection. Jacoby et al. noted that use of a generate–recognize strategy serves as an alternative means of excluding old words and that participants' reliance on such a strategy would violate assumptions underlying the equations used in their estimation procedure. Exclusion on the basis of a generate–recognize strategy refers to cases in which an old word automatically comes to mind as a completion for a stem, without being recollected, but is then subjected to a voluntary, recognition-memory check and withheld because it is recognized as old (cf. Jacoby & Hollingshead, 1990).

Suppose instructions for the inclusion test were changed such that participants were asked to complete stems with the first word that came to mind. This change would make the inclusion test equivalent to an indirect test of memory and, potentially, eliminate intentional use of memory—recollection. If performance on the inclusion test with changed instructions (indirect test) reflected only automatic influences of memory, the change would eliminate the effect of manipulations, such as full versus divided attention, that selectively influence recollection. Further, suppose that exclusion test instructions were changed by telling participants to use recognition memory to avoid completing stems with old words. This change in exclusion instructions would increase the likelihood of participants successfully excluding old words because recognition of a word as old is generally easier than is recalling the word. The change in instructions would also influence base-rate performance. Because of false recognition, words that were not earlier studied would sometimes be mistakenly excluded and, so, base rate for the exclusion test would be lower than for the inclusion test.

The changes in inclusion and exclusion instructions are such that the inclusion test would now measure the probability of a study word being generated as a completion, whereas the exclusion test would measure the success of a generate–recognize strategy as a means of excluding old words. Participants' reliance on a generate–recognize strategy would violate the assumptions underlying the equations used by Jacoby et al. (1993) to estimate R and A. Rather than R being equivalent for the inclusion and exclusion tests, recollection would not be used for either of the types of test and recognition would be important only for the exclusion test. The independence assumption would also be violated. A word must be generated before it can be recognized, and so conscious memory would not be independent of automatic influences of memory involved in generating a completion.

Paradoxical Dissociations and Correlations as Tests of Independence

Curran and Hintzman (1995) examined recall performance cued with word stems and obtained results that they interpreted as showing that participants' reliance on a generate–recognize strategy, along with correlation between processes at the level of items, invalidated the independence assumption underlying the process-dissociation procedure. They manipulated study duration and found what they termed a *paradoxical dissociation* between *R* and *A*. Increasing study time produced an increase in *R* but a decrease in *A*. Curran and Hintzman argued that this was because violation of the independence assumption resulted in *A* being underestimated by an amount that increased with the magnitude of *R*. The dissociation is paradoxical because experiments with indirect tests of memory have shown that manipulating study time leaves performance unchanged (Greene, 1986; Jacoby & Dallas, 1981). Consequently, increasing study time would be expected to increase *R* and leave *A* relatively invariant—the same form of process dissociation produced by manipulating full versus divided attention during study.

Significant correlations between *R* and *A* were found by Curran and Hintzman (1995) and treated as "direct evidence" of violation of the independence assumption underlying the process-dissociation procedure. They questioned whether assumptions were satisfied in earlier experiments with the process-dissociation procedure and advocated the use of correlations to directly test the independence assumption in future experiments. The Curran and Hintzman (1995) article was followed by debate about factors responsible for the paradoxical dissociation that they observed and their use of correlation to diagnose violation of the independence assumption (Curran & Hintzman, 1997; Hintzman & Curran, 1997; Jacoby, Begg, & Toth, 1997; Jacoby & Shrout, 1997). Jacoby and Shrout (1997) provided a psychometric analysis of effects of violations of independence on correlations between *R* and *A* and argued that the correlations reported by Curran and Hintzman did not speak to the independence assumption underlying the process-dissociation procedure. More is said about this when the results of Experiment 1 are reported.

Experiment 1

In Experiment 1, we manipulated instructions and examined the effects of full versus divided attention during study on stem-cued recall. Direct-retrieval instructions, similar to instructions used by Jacoby et al. (1993), were expected to produce results showing that dividing attention reduced *R* but left *A* relatively invariant, replicating results reported by Jacoby et al. Generate–recognize instructions, used in a second condition, were expected to produce a very different pattern of results. As compared with direct-retrieval instructions, generate–recognize instructions were expected to produce poorer performance on the inclusion test but to increase the accuracy of exclusion performance. Reliance on a generate–recognize strategy makes recollection irrelevant for the inclusion test but makes recognition-memory performance important for exclusion performance.

Base rates were set to be sufficiently high to avoid zero scores on the exclusion test in the direct-retrieval condition and were not expected to differ across inclusion–exclusion tests in that condition. (For a discussion of the importance of avoiding zero scores for the exclusion test, see Jacoby, Begg, & Toth, 1997, along with the response by Curran & Hintzman, 1997). However, for the generate–recognize condition, base rate was expected to be lower for the exclusion than for the inclusion test. Further, a paradoxical dissociation was expected in the generate–recognize condition—dividing attention was expected to decrease *R* but to increase *A*. Higher recognition-memory performance after full attention, as compared with divided attention, would produce a larger increase in accuracy in exclusion performance and, thereby, produce a larger artifactual decrease in *A*.

Correlations between *R* and *A* were computed to examine empirically Curran and Hintzman's (1995) claims about the utility of correlations for detecting violations of the independence assumption underlying the process-dissociation procedure. Jacoby and Shrout's (1997) psychometric analysis led us to expect significant correlations between *R* and *A* even in the direct-retrieval condition. It is not truly legitimate to examine correlations between *R* and *A* in the generate–recognize condition. If estimates are invalid because of violations of assumptions underlying the estimation procedure, correlations between the estimates are not meaningful. However, such correlations were computed for purposes of comparisons with those from Curran and Hintzman's Experiment 5, their only experiment using an inclusion–exclusion procedure that showed a paradoxical dissociation when zero scores for the exclusion test were removed. As is shown later, results from that experiment are similar to those from the generate–recognize condition.

Method

Participants. Ninety-six participants, 60 from the University of Texas at Austin and 36 from McMaster University, participated in the experiment in return for credit in an introductory psychology course. Half of the participants were randomly assigned to the direct-retrieval test condition, and the other half were assigned to the generate–recognize test condition. An additional 8 participants were tested, but their results were not used for purposes of analyses. Two of those participants were unable to do the divided-attention task, and the remaining 6 participants had zero scores in the exclusion task. Those having zero scores were all from the generate–recognize condition, and 5 of the 6 zero scores were for words studied under conditions of full attention. Participants were tested individually.

Materials and design. Words used in the experiment comprised a pool of 141 five-letter nouns of low, medium, and high frequency as indexed by Thorndike and Lorge (1944). Materials used in this experiment, along with those used in Experiments 2 and 3, appear in Appendix A. One hundred and twenty of these words were divided into three sets of 40 words each. Across formats, sets of words were rotated through the three study-presentation conditions: full attention, divided attention, and new (not studied). Each of the sets was divided further into two sets of 20 words each, which were rotated through the two test conditions: inclusion and exclusion. This arrangement resulted in 6 formats (3 presentation conditions × 2 test conditions). Subsets of words had an equal distribution of word frequency ($M = 34.7$, range = 33.6 to 36.2), set size (the number of five-letter word completions for the stem; $M = 3.8$, range = 3.7 to 3.9), and base rate. (On the basis of previous studies using the same materials, the probability of completing stems in each subset with the target solution when new had a mean of .44 and ranged from .437 to .448.)

To avoid primacy and recency effects, we presented 5 items at the beginning and another 5 items at the end of both the full-attention and divided-attention study lists. These buffer items stayed constant across all formats. This resulted in two study lists of 50 words each (40 critical and 10 buffer items). The divided-attention study list was always presented prior to the full-attention study list.

The test list consisted of 120 three-letter word stems corresponding to the 40 full-attention study words, 40 divided-attention study words, and 40 new words. Each of the 3-letter word stems was unique within the experiment but not within the language. That is, each stem could be completed with more than one 5-letter word but only one of the completions appeared within the experiment (e.g., *mer—*; *mercy, merge, merit,* and *merry*). For each word type (i.e., full attention, divided attention, and new), half of the word stems were presented in the inclusion-test condition, and half were presented in the exclusion-test condition. For the direct-retrieval test condition, the inclusion- and exclusion-test trials were intermixed to produce one test list of 120 items. For the generate–recognize test condition, the inclusion- and exclusion-test trials were blocked so that there were two 60-item test lists.[1] The exclusion test was always presented first. A short practice list of 5 items was presented prior to the beginning of each test list. For the direct-retrieval condition, the practice consisted of 3 inclusion-test items with 1 full-attention, 1 divided-attention, and 1 new test item; and 2 exclusion-test items with 1 full-attention and 1 divided-attention item. For the generate–recognize condition, there was a practice before each test block and these practice lists consisted of 1 full-attention item, 1 divided-attention item, and 3 new items. In all phases of the experiment, order of presentation was random with the restriction that not more than 3 items representing the same combination of conditions could be presented in a row.

The listening task used in the divided-attention condition was one previously used by Craik (1982). In this task, participants monitored a tape-recorded list of digits to detect target sequences of three odd (as opposed to even) numbers in a row (e.g., *9, 3, 7*). The digits were random, with the exception that a minimum of one number and a maximum of five numbers occurred between the end of one target sequence and the beginning of the next target sequence. Digits were recorded at a 1.5-s rate.

Procedure. Words were presented and responses were collected on a PC-compatible computer interfaced with a VGA-color monitor by using Schneider's (1990) Micro-Experimental Laboratory (MEL; Version 1.0) software system. The character size of the stimuli was approximately 3 × 5 mm. Words were presented in white letters on a black background in lowercase letters in the center of the screen.

For both study lists, words were presented on the computer screen, one word at a time. The words appeared for 1.5 s followed by 0.5 s of blank screen. For the full-attention list, participants were instructed to read the words aloud and to remember them for a later memory test. For the divided-attention list (presented first), participants were told that they were to do two tasks at the same time: a listening task and a reading task. They were informed that it was very important not to miss a target sequence in the listening task. Participants responded in the listening task by pressing a key whenever they detected a target sequence. They were informed that while doing the listening task, they would be presented with a list of words that they were to read aloud. However, they were cautioned not to allow the reading of the words to disrupt their performance on the listening task.

In the final phase of the experiment, word stems consisting of the initial three letters of a word followed by two dashes were presented one stem at a time on the computer screen. Direct-retrieval and generate–recognize instructions are presented in Appendix B. In the direct-retrieval test condition, each word stem was preceded by the presentation of either the prompt *old* or the prompt *new* centered two lines above the word stem in capital letters. The prompt was presented 500 ms prior to the presentation of the word stem and remained on the screen with the word stem until the participant responded or until the deadline of 15 s elapsed. Participants were told to use the word stems as cues for recall of words that had been presented in either of the study lists they had read. However, they were informed that recall of a previously presented word would not always be possible because some of the word stems could only be completed with words that had not been presented in the study lists. For a word stem presented with the prompt *old,* participants were told to use a recalled word as a completion. However, for a word stem presented with the prompt *new,* they were told not to use a recalled word as a completion. Rather, they were to complete the word stem with a word different from the recalled word. When unable to recall a studied word, they were told that they should complete the stem with the first 5-letter word that came to mind, regardless of whether the stem was accompanied by the prompt *old* or the prompt *new.* Although participants were told to complete as many word stems as possible, they were reminded that it was important to use recalled words to complete word stems accompanied by the prompt *old* and not to use recalled words to complete word stems accompanied by the prompt *new.* If they could not think of an alternative to a recalled word to complete a stem for an exclusion test, participants were told that they should leave the word stem incomplete.

For the generate–recognize test condition, word stems were presented without prompts, and the inclusion and exclusion test items were blocked to create separate test lists. Participants were informed that their task was to complete word stems and that some of the stems could only be completed by new words. In the exclusion test condition, the participants were informed that we were interested in seeing whether people could avoid using the earlier-presented study words as completion words. Therefore, they were to check each completion word that came to mind, before giving it as a response, to be certain that it was not an earlier-presented word. If the word seemed at all familiar, they were not to give it as a response but, rather, were to think of an alternative completion. If they could not think of an alternative completion, they were told to leave the stem incomplete and wait for the 15-s deadline to elapse. In the inclusion condition, the participants were informed that for this test, we were interested in seeing how quickly they could complete the stem without worrying about whether their completion words were presented earlier. They were told not to try to use memory because it would slow them down but to simply give the first 5-letter completion word that came to mind that fit the stem and to do so as rapidly as possible.

For all tests, participants were told that completion words were to be five letters long and that no plurals or proper names were

[1]Experiment 1 confounded the manipulation of instructions with a difference in test order. For the direct-retrieval condition, inclusion and exclusion tests were intermixed in the same way as done by Jacoby et al. (1993). For the generate–recognize condition, in contrast, the exclusion test preceded the inclusion test. A preliminary experiment combined generate–recognize instructions with intermixed tests. Results from the few participants tested in that experiment suggested that participants were hesitant to adopt a generate–recognize strategy. Large differences in base rate were not present, and *A* was relatively invariant across the manipulation of attention. Separating inclusion and exclusion tests was done to produce a more dramatic difference between the direct-retrieval and generate–recognize conditions. Experiment 3 intermixed inclusion and exclusion tests for both conditions so that the only difference between conditions was created by manipulating instructions.

allowed. If the participant's response met these criteria, we pressed a key to remove the word stem from the screen and then pressed another key to present the next trial. Otherwise, participants were informed of their error and were told to attempt to give a satisfactory completion for the word stem. If the word stem had not been completed after the allotted time, a beep sounded, the screen cleared, and we initiated the next trial.

Statistical analyses. There were two main sets of analyses carried out on the data for this experiment. For each participant, we calculated the proportion of stems completed with the target solution in each Instruction × Test × Study Condition. We performed ANOVAs on these data as well as on estimates of R and A derived from these data. To compute correlations, we calculated estimates of A for new words in the same way as done by Curran and Hintzman (1995). Base rate for the exclusion test was subtracted from base rate for the inclusion test to estimate the probability of false recollection (FR). The base rate for the exclusion test was then divided by $1 - FR$ to estimate A for new items. For all ANOVAs and power analyses, alpha was set to .05 unless otherwise noted.

Results and Discussion

In the divided-attention condition, the probability of failing to detect a target sequence for the listening task was .12 for the direct-retrieval condition and .13 for the generate–recognize condition.

Proportion of stems completed with old words. Performance on the inclusion test (Table 1) was higher in the direct-retrieval test condition than in the generate–recognize test condition, $F(1, 94) = 5.58$, $MSE = 0.019$, and was higher after full attention than after divided attention to study, $F(1, 94) = 16.41$, $MSE = 0.009$. Although the interaction of instruction and attention was not significant, the advantage produced by full attention during study was numerically larger in the direct-retrieval condition than in the generate–recognize condition. For the exclusion test, participants were less likely to mistakenly use an old word as a completion in the generate–recognize test condition, as compared with direct retrieval, $F(1, 94) = 64.83$, $MSE = 0.033$. Exclusion performance was also more accurate after full attention to study, as compared with divided attention, $F(1, 94) = 35.79$, $MSE = 0.013$. In the direct-retrieval condition, the difference in baseline completion rates for the inclusion and exclusion tests did not approach significance, $F < 1$. However, in the generate–recognize condition, base rate was much lower for the exclusion test than for the inclusion test, $F(1, 47) = 43.43$, $MSE = 0.013$. The overall pattern of results shows the expected differences between direct-retrieval and generate–recognize strategies.

Estimates of R and A. The probability of recollection (Table 2) was higher in the generate–recognize test condition than in the direct-retrieval test condition, $F(1, 94) = 26.05$, $MSE = 0.049$. This result was expected because for the generate–recognize condition, R does not truly measure recollection but, rather, measures recognition memory of words that were generated as a completion. For both test conditions, R was higher after full attention than after divided attention to study, $F(1, 94) = 44.85$, $MSE = 0.025$.

The analysis of A revealed a significant interaction between study and test conditions, $F(1, 94) = 7.87$, $MSE = 0.012$. Results from the direct-retrieval condition showed a process dissociation that replicated results reported by Jacoby et al. (1993, Experiment 1b). Dividing attention during study produced a decrease in R, $F(1, 47) = 19.57$, $MSE = 0.029$, but left A almost perfectly invariant, $F < 1$. This result did not reflect insensitivity of our measure because the power to detect an effect on A as large as that observed in the generate–recognize condition was .99 (Cohen's $d = 0.83$). Estimates of A for old items were significantly above base rate, $F(1, 47) = 32.24$, $MSE = 0.008$,

Table 1
Proportion of Stems Completed With a Target Word (From Participant Means)

Test condition and experiment	Inclusion			Exclusion		
	Full/10 s	Div/1 s	New	Full/10 s	Div/1 s	New
Direct retrieval						
Current experiments						
Experiment 1	.70	.62	.45	.40	.48	.43
Experiment 2	.72	.64	.38	.33	.45	.36
Experiment 3	.77	.66	.41	.34	.45	.39
Curran & Hintzman (1995)						
Experiment 4	.60	.49	.31	.25	.31	.31
Generate–recognize						
Current experiments						
Experiment 1	.63	.59	.45	.16	.29	.29
Experiment 2[a]	.72	.64	.38	.07	.20	.27
Experiment 3	.71	.67	.46	.19	.33	.37
Curran & Hintzman (1995)						
Experiment 5	.59	.55	.37	.12	.22	.31
Experiment 3[b]	.63	.54	.38	.13	.22	.26

Note. Study conditions for Experiment 1 were full and divided (Div) attention; study conditions for Experiments 2 and 3 and for Curran & Hintzman (1995) were presentation durations of 10 s and 1 s. [a]Data points simulated from remember–know data. [b]Data points simulated from recollect and exclude data.

Table 2
Estimates of Recollection and Automaticity (From Participant Means)

Test condition and experiment	Recollection (*R*) Full/10 s	Recollection (*R*) Div/1 s	Automaticity (*A*) Full/10 s	Automaticity (*A*) Div/1 s
Direct retrieval				
Current experiments				
Experiment 1	.29	.14	.54	.55
Experiment 2	.39	.19	.52	.54
Experiment 3	.44	.22	.59	.58
Curran & Hintzman (1995)				
Experiment 4	.32	.16	.40	.39
Generate–recognize				
Current experiments				
Experiment 1	.46	.30	.30	.40
Experiment 2	.65	.44	.19	.35
Experiment 3	.52	.34	.39	.50
Curran & Hintzman (1995)				
Experiment 5	.40	.30	.28	.37
Experiment 3	.48	.30	.27	.34

Note. Study conditions for Experiment 1 were full and divided (Div) attention; study conditions for Experiments 2 and 3 and for Curran & Hintzman (1995) were presentation durations of 10 s and 1 s.

showing a large effect of study on automatic influences of memory.

A paradoxical dissociation was found in the generate–recognize condition. Dividing attention decreased *R*, $F(1, 47) = 26.18$, $MSE = 0.022$, and had the opposite effect of increasing *A*, $F(1, 47) = 20.94$, $MSE = 0.010$. For words studied under full attention, *A* was significantly below baseline, $F(1, 47) = 13.26$, $MSE = 0.008$, which, as noted by Curran and Hintzman (1995), is a certain sign that the assumptions underlying the estimation procedure were violated. Participants' reliance on a generate–recognize strategy produced a paradoxical dissociation by violating both the independence assumption and the assumption of equality of *R* for inclusion and exclusion tests.

The results of Experiment 1 showed that the process dissociation reported by Jacoby et al. (1993) is replicable when direct-retrieval instructions are used. Using similar instructions, Schmitter-Edgecombe (1996, Experiment 1) independently replicated results reported by Jacoby et al. Her results showed that *A* was unchanged by full versus divided attention (.25 and .25), although *R* was reduced by dividing attention (.28 vs. .09). Base rates were the same for inclusion and exclusion tests (.17 and .17). Of the 32 participants in her experiment, 3 participants produced zero old words in the exclusion condition. Removing their data left *A* relatively unchanged by full versus divided attention (.27 and .26).

Instructions are important for satisfying assumptions underlying the estimation procedure. In contrast to results found by using direct-retrieval instructions, generate–recognize instructions produced a paradoxical dissociation. Significant differences in base rate between inclusion and exclusion tests provide direct evidence of the violation of assumptions. Such differences were found when generate–recognize instructions were used but were not found with direct-retrieval instructions. Next, we turn to the question of whether correlations between *R* and *A* can be used to directly test assumptions underlying the process-dissociation procedure.

Violations of the independence assumption: Diagnosticity of correlations. Curran and Hintzman (1995) reported correlations between *R* and *A* that they interpreted as direct evidence that the independence assumption underlying the process-dissociation procedure had been violated. In response, Jacoby and Shrout (1997) provided a psychometric analysis of effects of violations of independence on correlations between *R* and *A*. That analysis distinguishes between *process dependence* and *aggregation bias*. Process dependence results when participants rely on a strategy that makes conscious memory dependent on automatic influences of memory, such as a generate–recognize strategy. Aggregation bias can result when parameters are estimated by aggregating across participant or item data, and estimates are correlated at the level across which data were aggregated. Both process dependence and aggregation bias reflect a correlation that cannot be directly observed but only imagined. This is true because it is necessary to aggregate across something, either participants or items, to compute correlations.

In a postscript, Jacoby and Shrout (1997) summarized the exchange with Curran and Hintzman (Curran & Hintzman, 1997; Hintzman & Curran, 1997; Jacoby, Begg, & Toth, 1997) as having reached agreement that observed correlations of *R* and *A* calculated by aggregating over participants or items cannot be used to provide evidence of process dependence. However, disagreement about the use of those correlations as evidence for aggregation bias was unresolved.

Correlation at a particular level will bias estimates only if one aggregates across the level at which the correlation exists before computing estimates. As a commonplace example of why aggregation is necessary to bias estimates, height and weight are correlated across people. However, that correlation does not mean that when measuring the height or weight of an individual, one has to worry that the measure on the one dimension is biased by the value on the other dimension. Similarly, we estimated *R* and *A* for each participant, and so, correlation at the level of participants is not a source of bias for estimates.

Curran and Hintzman (1995) used the observed, positive correlation of *R* and *A* at the item level to infer that the unobservable correlation at the participant–item level, across which scores are aggregated to compute participant-level estimates, was also positive. However, Jacoby and Shrout (1997) argued that one cannot make clear inferences about the unobservable correlation responsible for aggregation bias on the basis of correlations observed at the item or participant level. This is because correlations are between estimates, and one source of correlations is the estimation procedure itself. Recall that the estimates make use of probabilities of reporting an old word in the inclusion-test and the exclusion-test conditions. The nonlinear dependence of estimates of *R* and *A* on the same two empirical facts induces correlation between them. A mathematical analysis showed that the correlation produced by the estimation procedure can be either positive or negative, depending on the underlying parameters for the automatic and recollective memory processes (Jacoby & Shrout, 1997).

Curran and Hintzman (1995) computed estimates of R and A for base-rate items (new) as well as for old items. Estimates of R for old items (R_{old}) and A for new items (A_{new}) come from different trials so correlations between them are not based on the same empirical facts and cannot be a result of the estimation procedure. This makes it useful to compare correlations between R_{old} and A for old items (A_{old}) with those between R_{old} and A_{new} to examine the contribution to correlation that might come from the estimation procedure.

Correlation at the level of participants. Curran and Hintzman (1995) interpreted their finding of an inverse correlation between R and A at the level of participants as providing evidence that participants relied on a generate–recognize strategy. We computed estimates of R_{old} and A_{old} by aggregating across items and conditions of attention for each participant. The correlation between R_{old} and A_{old} at the level of participants (Table 3) was significant in the direct-retrieval condition but was not significant in the generate–recognize condition. This result demonstrates that use of a generate–recognize strategy does not always produce an inverse correlation between R and A at the level of participants. Also, it cannot be argued that the significant correlation between R_{old} and A_{old} in the direct-retrieval condition proves that participants in that condition relied on a generate–recognize strategy. The significant correlation may have come from other sources such as the estimation procedure itself. In line with that possibility, the correlation between R_{old} and A_{new} was not significant in the direct-retrieval condition.

Correlation at the level of items. Findings of significant correlations at the level of items were interpreted by Curran and Hintzman (1995) as direct evidence of violation of the independence assumption underlying the process-dissociation procedure. To compute those correlations (Table 4), we aggregated across participants and conditions of attention. Correlations obtained by Curran and Hintzman (1995) in their Experiment 5 are presented for purposes of comparison. The correlation between R_{old} and A_{old} was significant in both the direct-retrieval and the generate–recognize conditions. Curran and Hintzman (1995) found that the correlation between R_{old} and A_{new} was as high as that between R_{old} and A_{old}, which they (Hintzman & Curran, 1997) took as strong evidence that item differences, rather than the estimation procedure, were responsible for both correlations. As shown in Table 4, we also found the correlation of R_{old} and A_{new} to be as high as that between R_{old} and A_{old} but only in the generate–recognize condition. In the direct-retrieval condition, the correlation between R_{old} and A_{new}, computed on those same items, was near zero. For the direct-retrieval condition, one cannot dismiss the possibility that the correlation between R_{old} and A_{old} was produced by the estimation procedure.

One concern is that correlations at the item level might unduly reflect the contribution of items that gave rise to

Table 3
Correlations Between Estimates of R and A Calculated From Participant Means

Condition and experiment	$R_{old} - A_{new}$	$R_{old} - A_{old}$
Direct retrieval		
Current experiments		
Experiment 1	.04	−.43**
Experiment 2	.08	−.52*
Jacoby & Hay (in press)	−.03	−.59*
Generate–recognize		
Current experiments		
Experiment 1	−.33*	−.19
Experiment 2	.07	−.21
Jacoby & Hay (in press)	−.31	−.66**
Curran & Hintzman (1995)		
Experiment 5	−.06	−.18
Experiment 3	−.05	−.35*

Note. R = the estimate of the probability that participants could complete a stem with an old word because they recollected the studied word; A = the estimate of the probability that participants could complete a stem with an old word because the old word came automatically to mind.
$^{*}p < .05$. $^{**}p < .01$.

perfect performance on either the inclusion test or the exclusion test. Indeed, eliminating items that produced either a probability of 1.0 for inclusion or a probability of 0 for exclusion reduced the correlation between R_{old} and A_{old} for the direct-retrieval condition but not for the generate–recognize condition.

Comparison of item-based and participant-based estimates. Curran and Hintzman (1995) used the positive correlation between R and A at the item level to infer that the unobservable correlation responsible for the bias produced by aggregating across items for each participant was also positive. Because of the unobservable positive correlation, A was said to be *underestimated* by an amount that increased with increases in R, producing a paradoxical dissociation. Their arguments can also be applied to predict results for item-based estimates. There, correlation at the item level cannot be a source of bias because estimates were obtained for each item. Rather, the potential source of aggregation bias is produced by aggregating across participants to gain estimates of R and A for each item. One can use the observed correlation at the participant level to infer the unobservable correlation responsible for bias produced by aggregating across participants, just as did Curran and Hintzman to infer the correlation at the unobservable participant–item level. However, the correlation at the participant level is negative, which means that the same should be true for the unobservable correlation at the item–participant level, and that A should be *overestimated* by an amount that increases with increases in R.

For the direct-retrieval condition, the inverse correlation between R and A at the level of participants (−.43) was of similar magnitude, but opposite in direction, to the correlation at the level of items (.38). Consequently, one might expect paradoxical dissociations of similar magnitude but opposite forms for item-based and participant-based estimates. However, the pattern of results for the direct-retrieval

Table 4
Correlations Between Estimates of R and A Calculated From Items Means

Condition and experiment	$R_{old} - A_{new}$	$R_{old} - A_{old}$	n
Direct retrieval			
Current experiments			
Experiment 1	−.02	.38** (.22*)	120 (83)
Experiment 2	.19	.38** (.34*)	96 (46)
Jacoby & Hay (in press)	.14	.33** (.25)	120 (58)
Generate–recognize			
Current experiments			
Experiment 1	.31**	.38** (.38**)	120 (63)
Experiment 2	.48**	.55** (.46*)	96 (20)
Jacoby & Hay (in press)	.33**	.37** (.29*)	120 (58)
Curran & Hintzman (1995)			
Experiment 5	.57**	.55**	
Experiment 3	.70**	.70**	

Note. Numbers in parentheses are the correlations, with corresponding number of observations, calculated after removing any item that had an A estimate that was undefined, 1, or 0. R = the estimate of the probability that participants could complete a stem with an old word because they recollected the studied word; A = the estimate of the probability that participants could complete a stem with an old word because the word came automatically to mind. n = number of observations.
*$p < 05$. **$p < .01$.

condition was the same whether estimates were computed for each participant or for each item. When items that produced perfect scores on inclusion or exclusion tests were eliminated, estimates computed for each item showed that dividing attention reduced R (.22 vs. .11) but left A relatively invariant (.50 vs. .49). Base rate for those items was .40. Similarly, when computed for each participant (none of whom had perfect scores on inclusion or exclusion tests) estimates of R were reduced by dividing attention (.29 vs. .14), but estimates of A were relatively unchanged (.54 vs. .55). Base rate for those participants was .44.

The comparison of results computed from participant means with those computed from item means is a variant of the strategy of using a quasi-F ratio to avoid the "language-as-fixed effect fallacy" (Clark, 1973). How is it possible to find significant correlations at the participant level and at the item level and not have bias result from aggregating across either of the two levels to find estimates? The correlation that is important for aggregation bias is not at the participant level or at the item level but, rather, at the item–participant level or participant–item level, dependent on how estimates are computed. The correlations responsible for aggregation bias cannot be directly observed, nor can they be inferred from the correlations observed at the participant or item level (Jacoby & Shrout, 1997).

Although it is possible for aggregation bias to dramatically distort estimates (Hintzman & Curran, 1997), it seems unlikely that aggregation bias played a role in producing results observed for the direct-retrieval condition. To argue otherwise, one has to claim that the paradoxical dissociation that should have been, but was not, observed when estimates were obtained for each participant was offset by a true effect of dividing attention on A that was opposite to that which offset the paradoxical dissociation that should have been, but was not, observed when estimates were obtained for each item. A paradoxical dissociation was obtained in the generate–recognize condition. However, that paradoxical dissociation was produced by violating assumptions underlying the estimation procedure rather than by aggregation bias.

Instructions are an important boundary condition for meeting assumptions underlying the process-dissociation procedure. Results of Experiment 1 showed that direct-retrieval instructions produced effects that replicated those reported by Jacoby et al. (1993). However, participants' reliance on a generate–recognize strategy violates assumptions of the estimation procedure (Curran & Hintzman, 1995; Jacoby et al., 1993). The difference in base rates for inclusion and exclusion tests indexed participants' reliance on a generate–recognize strategy—base rate was significantly lower for the exclusion test only when generate–recognize instructions were used. Correlations, in contrast, were not useful for testing whether assumptions of the estimation procedure were met. Observed correlations in the direct-retrieval condition might have come from sources other than violation of the independence assumption, including the estimation procedure itself.

Experiment 2

Reliance on subjective reports provides an alternative to the process-dissociation procedure as a means of separating the contributions of recollection and automatic influences of memory. Following Tulving (1985), Gardiner and his colleagues (e.g., Gardiner, 1988; Gardiner & Java, 1991; Gardiner & Parkin, 1990) have used a remember–know procedure to investigate the phenomenology that accompanies memory performance. Participants are instructed to respond "remember" only if they can remember the details surrounding the study presentation of a test item and to respond "know" if they feel certain that a test item was earlier studied but are unable to recall the details of its study presentation.

In Experiment 2, we combined the remember–know procedure with the independence assumption from the process-dissociation procedure to examine effects of study time in recall cued with word stems. Just as was done with instructions for an inclusion test, participants were told to use stems as a cue for recall of an earlier-studied word or, if they could not do so, to complete stems with the first word that came to mind. However, participants were also required to report on the subjective experience that accompanied their production of a completion word. Immediately after producing a completion word, participants were to classify the completion word as one that they "remember" as earlier studied, one that they "know" was earlier studied, or as "new," not earlier studied.

For our independence/remember–know (IRK) procedure, the probability of "remembering" served as a measure of recollection (e.g., Jacoby, Yonelinas, & Jennings, 1997; Lindsay & Kelley, 1996; Yonelinas & Jacoby, 1996). Participants should classify an old word as "know" or "new" only if the word came automatically to mind but was not recollected as earlier studied: $A(1 - R)$. That combination is the same as for mistakenly producing an old word on an exclusion test. Consequently, the independence assumption can be used to estimate A as $[P(\text{Know}) + P(\text{New})]/[1 - P(\text{Remember})]$.

In Experiment 1, instructing participants to use a generate–recognize strategy produced a paradoxical dissociation. Reliance on that strategy resulted in participants excluding words that came to mind automatically and were then recognized as well as words that were recollected. Recognition without recollection describes words judged as "know" in the IRK procedure. When using a generate–recognize strategy for an exclusion test, only old words that would be judged as "new" should be mistakenly used as a completion. To mimic effects of a generate–recognize strategy, words that participants "know" were earlier studied can be grouped with old words judged as "remember" rather than with those judged as "new" when estimating A: $P(\text{New})/1 - [P(\text{Remember}) + P(\text{Know})]$. Estimating A in this way was expected to produce a paradoxical dissociation that was the same as that found in the generate–recognize condition in Experiment 1.

The IRK procedure is a refinement of a procedure that Curran and Hintzman (1995) used to produce a paradoxical dissociation by manipulating study time. For their *recollect and exclude* procedure, participants tried to give two responses to each stem. The probability of recollection was measured as the probability of an old word being written in a column labeled *remember*. In contrast, the probability of writing an old word in a column labeled *new* was said to be equivalent to the probability of mistakenly using an old word as a completion for an exclusion test and, consequently, was divided by $[1 - P(\text{Remember})]$ to estimate A. This division is motivated by the assumption that R and A independently contribute to performance. Curran and Hintzman later rejected their new procedure because it too showed violation of the independence assumption by revealing a paradoxical dissociation produced by the manipulation of study time.

A weakness of the recollect and exclude procedure is that participants are not allowed to distinguish between words that they "remember" and words that they only "know" were studied earlier. Consequently, it seems likely that participants would, at least sometimes, write words that they only "know" are old in the "remember" column. Their doing so would result in an overestimate of recollection along with a paradoxical dissociation just as would the generate–recognize version of our IRK procedure or use of a generate–recognize strategy in combination with the inclusion–exclusion procedure. Results from the generate–recognize version of the IRK procedure were compared with results from Curran and Hinzman's (1995) recollect and exclude procedure to show that both the paradoxical dissociation and correlations between R and A were the same for the two procedures. The probability of false recollection, measured by the probability of responding "remember" to new words, was examined to assess whether false recollection was likely to have played a role in producing base-rate differences observed with the inclusion–exclusion procedure.

Method

Participants. Twenty-four students participated in the experiment in return for credit in an introductory psychology course at McMaster University. One additional participant was tested, but his data were discarded because he did not classify any recalled words as "know," suggesting that he failed to understand instructions.

Materials and design. To accommodate the change from the inclusion–exclusion procedure in Experiment 1 to the remember–know procedure in this experiment, we modified the materials and design slightly. The only changes in the materials were in the size of sets used to construct lists, which changed from 40 items per set to 32 items per set, and in the overall base rate of the items when new, which changed from .44 to .35. Rather than manipulating attention during study, presentation duration (1 s vs. 10 s) for study was manipulated, as was done in Curran and Hintzman's (1995) experiments. Two study lists of 42 items each (32 critical items and 10 buffer items) were presented, with items in one list presented for 10 s each and items in the other list presented for 1 s each. The order of presentation of the study lists was balanced across participants so that half of the participants had the long study-duration list first and the other half had it second. There was only one test list that comprised 96 items; 32 stems represented each of the three types of word (10-s study, 1-s study, and new).

Procedure. For both study lists, participants saw a list of words on the screen presented one word at a time. Participants were instructed to read the words aloud and to try to remember them for a later memory test. After the first list, participants were simply reminded prior to the presentation of the second list to read the words and to try to remember them for the following memory test.

Instructions for the test appear in Appendix B. Participants were told to use the stem as a cue to help them recall a word that was presented in either of the study lists and to use the recalled word to complete the stem. If they could not recall a suitable study word, they were to complete the stem with the first 5-letter word that came to mind that fit the stem. Participants were told that no proper names or plurals were allowed as completions. Once a stem was completed, we pressed a key to clear the screen, and then the participant judged the completion word as "remember," "know," or "new." A "remember" response was to be given if the participants could consciously remember details of the prior encounter with the word in the study list. For example, they may remember some specific detail about its prior presentation, such as an image, or some personal significance they may have thought of at the time of study, or they may remember how the word looked on the screen. "Know" meant participants knew for a fact that the word was presented in study but could not remember any specific details about its prior presentation. "New" meant the word had not been presented earlier in either of the study lists. Once we had entered each participant's decision, the next trial was presented after a 0.5-s delay. If the word stem was not completed within the alloted time, a beep sounded, and we initiated the next trial.

Results and Discussion

Proportion of stems completed with old words. Performance on the stem-completion task was higher for old words than for new words, $F(1, 23) = 379.10$, $MSE = 0.002$, and words presented at the 10-s study duration were significantly more likely to be given as a completion than were words presented at the 1-s study duration, $F(1, 23) = 13.55$, $MSE = 0.007$. The probabilities of "remember," "know," and "new" responses, as a function of study duration, are shown in Table 5. Those probabilities were used to simulate performance on inclusion and exclusion tests (Table 1). Inclusion performance was estimated as the probability of producing an old word as a completion. To mimic direct retrieval, "know" responses were added to "new" responses to represent exclusion performance, whereas to mimic generate–recognize, only "new" responses represented exclusion performance. For direct retrieval, base rates were estimated as .38 for the inclusion test and .36 for the exclusion test. In contrast, for generate–recognize performance, new words that participants mistakenly classified as "remember" or "know" would be excluded, producing a base-rate difference between inclusion and exclusion tests (.38 vs. .27).

Use of the remember–know procedure allowed us to examine false recollection of new words. Results showed that the probability of mistakenly saying "remember" after producing a completion word that was actually new was very low (.02). Probably by coincidence, that probability was the same as the difference in base rates for inclusion and exclusion tests found for direct-retrieval conditions in Experiment 1. The slightly lower base rate for the exclusion test in that experiment might reflect false recollection. However, if it does, the probability of false recollection was not sufficient to produce a significant difference in base rates. In Curran and Hintzman's (1995) Experiment 3, which used the recollect and exclude procedure, they found the probability of "false recall" for new items to be .12—a probability nearly identical to the sum of false remembering and false knowing in our experiment (.11). Such a high probability of false recall makes it almost certain that participants in Curran and Hintzman's experiment misclassified as "remembered" words that would be called "know" if they had the option. Further, these results suggest that base-rate differences between inclusion and exclusion tests in the generate–recognize condition of Experiment 1 were largely produced by false "knowing" rather than by false recollection.

Estimates of R and A. Results using the direct-retrieval equations (Table 2), grouping "know" responses with "new" responses to compute estimates, showed that reducing study time from a 10-s to a 1-s duration decreased R, $F(1, 23) = 49.84$, $MSE = 0.010$, but left A almost perfectly invariant, $F < 1$. This result did not reflect insensitivity of our measure. The power to detect an effect on A as that observed when A was calculated in a way meant to mimic use of a generate–recognize strategy was >.995 (Cohen's $d = 1.35$). This result supports our prediction that A would not differ for the 10-s and 1-s duration conditions. Estimates of A for old items were significantly above baseline, $F(1, 23) = 33.17$, $MSE = 0.009$, showing a large effect of study on automatic influences of memory.

A paradoxical dissociation was found when "know" responses were grouped with "remember" rather than with "new" responses, the generate–recognize version of equations. Reducing study time from 10 s to 1 s decreased recollection, $F(1, 23) = 44.15$, $MSE = 0.012$, but had the opposite effect of increasing A, $F(1, 23) = 29.41$, $MSE = 0.009$. For words studied at the 10-s duration, A was significantly below baseline, $F(1, 23) = 42.93$, $MSE = 0.009$, which is a certain sign that the assumptions underlying the estimation procedure were violated.

The pattern of results for the generate–recognize version of the IRK procedure was the same as found in Curran and Hintzman's (1995) Experiment 3, which used the recollect and exclude procedure (Table 2). That similarity suggests that their paradoxical dissociation resulted because words that would have been called "know," if participants had the option of doing so, were sometimes misclassified as "remember." The difference in results is as much because of differences in "inclusion" performance as differences in "exclusion" performance (Table 1) and can be explained as being a result of recollection being higher in our experiment. The difference in recollection seems particularly pronounced for words in the 10-s study condition. For those items, inclusion performance is substantially higher, and exclusion performance is more accurate for the IRK procedure than for the recollect and exclude procedure. For their Experiment 3, Curran and Hintzman tested participants in groups of 1 to 10. Our procedure, in contrast, tested participants individually and required them to pronounce study words aloud. These differences in procedure are likely to explain the higher recollection in our experiment.

Correlations. The pattern of correlations was the same as observed in Experiment 1, but correlations were somewhat larger in magnitude. Although the number of participants in Experiment 2 was only half that of Experiment 1, the number of observations on which estimates were based was the same because a single test was used, rather than

Table 5
Mean Probabilities of Completion With Study Word and "Remember," "Know," and "New" Responses for Experiment 2

Condition	Remember	Know	New
Long (10 s)	.39	.26	.07
Short (1 s)	.19	.25	.20
New	.02	.09	.27

inclusion and exclusion tests. The correlation between R_{old} and A_{old} at the participant level (Table 3) was significant in direct-retrieval estimates, and the correlation at the item level (Table 4) was significant in both direct-retrieval and generate–recognize estimates. The correlation between R_{old} and A_{new} at the item level was significant only in generate–recognize estimates. The pattern of correlations found with generate–recognize estimates is strikingly similar to that found by Curran and Hintzman (1995) in their Experiments 3 and 5 (Table 4).

As in Experiment 1, direct-retrieval estimates revealed correlations at the participant and item levels that were approximately equal in magnitude but opposite in direction. If those observed correlations reflect the unobservable correlations responsible for aggregation bias, item-based estimates and participant-based estimates of R and A should show opposite paradoxical dissociations. However, the pattern of results for item-based estimates was the same as for participant-based estimates. Item-based estimates showed that increasing study duration increased R (.15 vs. .32) but left A unchanged (.54 vs. .54). Base rate for item-based estimates was .39.

For direct-retrieval estimates, it seems unlikely that aggregation bias played a role in the results. As stated before, to argue otherwise, one has to claim that the paradoxical dissociation that should have been, but was not, observed when estimates were obtained for each participant was offset by a true effect on A of increasing study time that was opposite to that which offset the paradoxical dissociation that should have been, but was not, observed when estimates were obtained for each item.

Results from other experiments using the IRK procedure with cued recall. Jacoby and Hay (in press) used the IRK procedure to examine the effect of full versus divided attention during study. The materials and procedure were the same as in Experiment 1, except that Jacoby and Hay used the remember–know test procedure of Experiment 2. The direct-retrieval version of the IRK procedure, using participant-based means, showed that divided attention during study as compared with full attention reduced R (.25 and .09) but left A unchanged (.61 and .61). Base rate was .46, and the probabilities of false recollection and false knowing were .02 and .11. The pattern of results was the same when item-based means were used—full versus divided attention left A relatively unchanged (.58 and .56). In contrast, mimicking a generate–recognize strategy by treating the sum of "remember" and "know" responses as a measure of recollection revealed a paradoxical dissociation. The pattern of correlations was the same as reported here (see Tables 3 and 4). Again, a significant correlation between R_{old} and A_{new} at the item level was found only with the generate–recognize version of the IRK procedure. Significant correlations that were observed cannot be used to test the independence assumption underlying the estimation procedure because of possible contributions to those correlations from other sources, including the estimation procedure itself.

The IRK procedure produced results that are the same as those found with the inclusion–exclusion procedure. Jacoby, Yonelinas, & Jennings (1997) also showed that values of estimates were extraordinarily close across the two procedures in a series of experiments done to examine cross-modality transfer in recall cued with word fragments. For "remember" responses to serve as a valid measure of recollection, participants must be aware of recollecting old words that come to mind as a completion for a stem. Such awareness is also required to use recollection as a means to avoid mistakenly producing old words on an exclusion test. The inclusion–exclusion procedure differs from the IRK procedure by requiring that participants use awareness as a basis for conscious control of responding, rather than only report on awareness. When adequate time is given for responding, the two procedures are likely to produce the same pattern of results. However, because the inclusion–exclusion procedure measures R as that which affords control over responses, and the IRK procedure measures phenomenological experience, we expect that the two need not always coincide.

Comparisons with results from other experiments using subjective report procedures. Mäntylä (1993) used a cued-recall procedure and found that age-related differences in memory influenced "remember" responses but left "know" responses unchanged. The same pattern of results was produced by manipulating study time. Our results agree with those of Mäntylä. If one looks only at "remember" and "know" responses, as he did, decreasing study time reduced the probability of a "remember" response but did not change the probability of a "know" response (Table 5). However, Mäntylä did not encourage participants in his experiments to guess, whereas we required participants to always produce a completion, guessing if necessary. By our view, knowing is on a continuum that includes items that participants call "new." Old items that are produced as a response and called "know" or "new" are combined to estimate automatic influences. Rather than try to eliminate guessing, we required guessing and separated automatic influences of memory from recollection (Jacoby, Yonelinas, & Jennings, 1997). Doing so shows that increasing study time increases recollection but leaves automatic influences of memory unchanged in recall cued with word stems.

Toth, Reingold, and Jacoby (1994) used the inclusion–exclusion procedure and found that manipulating level of processing influenced R but left A relatively unchanged. Experiments with stem-completion performance as an indirect test of memory have found small, if any, effect of manipulating level of processing (for a review, see Challis & Brodbeck, 1992). In contrast, Richardson-Klavehn and Gardiner (1996; Richardson-Klavehn, Gardiner, & Java, 1996) used a procedure that was very similar to Curran and Hintzman's (1995) recollect and exclude procedure and found a paradoxical dissociation produced by level of processing. Deep processing, as compared with shallow processing, was found to increase R but to decrease A. Interpretation of their results is made difficult by floor effects. Particularly after deep processing, many participants had zero scores for exclusion performance. Jacoby, Begg, & Toth (1997) described how such floor effects could produce a "paradoxical" dissociation. The interpretation of results is further complicated by the possibility that participants classified as "remember" completions that were produced because of automatic influences of memory but were then recognized as old. The problem is the same as described for Curran and Hintzman's recollect and exclude procedure and also applies to experiments done by De Houwer (in press) that used subjective reports to measure recollection. Richardson-Klavehn and Gardiner interpreted their results as reflecting "involuntary conscious memory," which refers to cases in which an old word involuntarily comes to mind as a completion followed by awareness that the word is old. Reingold and Toth (1996) discussed the relation between involuntary conscious memory and recognition memory in a generate–recognize strategy. The two notions seem very similar with the major difference being that the recognition in a generate–recognize strategy is not involuntary. We return to this topic in the General Discussion.

Experiment 3

Experiment 3 was similar to Experiment 1 but examined effects of study time rather than effects of full versus divided attention. Curran and Hintzman (1995) reported three experiments that used our inclusion–exclusion procedure to investigate the effects of presenting words for 1 s versus 10 s for study. In their first experiment, base rate was .12, and estimates computed by using results that included participants who performed perfectly on the exclusion test (0 scores) revealed a paradoxical dissociation by showing that A decreased with increased study time (.16 vs. .12). When zeros were removed, A was nearly identical for the two conditions (.18 vs. .17), and the difference was no longer significant. In their Experiment 4, they used stems having a higher base rate (.31) so as to reduce the likelihood of zeros for the exclusion test. Results from that experiment showed that A was near identical for the long-duration and short-duration conditions (.35 vs. .36). When zeros were removed, the small difference between conditions was reversed (.40 vs. .39). (See Jacoby, Begg, & Toth, 1997, and the response by Curran & Hintzmann, 1997, for a discussion of whether the invariance in A was produced by participants' misunderstanding of exclusion instructions.) In their Experiment 5, Curran and Hintzman (1995) used a practice session that was meant to ensure that participants properly understood exclusion instructions and found a paradoxical dissociation even when zeros in exclusion were removed. For that experiment, base rate for the exclusion test was significantly below that on the inclusion test. This difference in base rate was ignored when computing estimates.

Curran and Hintzman (1995) justified ignoring the significant difference in base rates by saying they were unable to ascertain any reason why an influence on estimates of A would result from doing so (p. 542). Further, they argued that the lower base rate for the exclusion test may have been produced by participants falsely recollecting some new items as old, and if so, the difference in base rates could be safely ignored. However, the low probability of false recollection observed in Experiment 2, reported here, makes it unlikely that false recollection was responsible for the significant base-rate difference observed in Curran and Hintzman's experiment. Experiment 3 was done to show that significant differences in base rate between inclusion and exclusion tests, comparable to those found by Curran and Hintzman, are important for finding a paradoxical dissociation. For a generate–recognize condition, we expected to find such differences in base rate along with a paradoxical dissociation just as was found in our Experiments 1 and 2. For a direct-retrieval condition, manipulating study time was expected to influence R but to leave A relatively invariant.

In Experiment 3, we used a base rate that was even higher than that used in Curran and Hintzman's (1995) Experiments 4 and 5. We expected the higher base rate to allow zero scores to be avoided in the direct-retrieval condition, although some zero scores might remain in the generate–recognize condition. Another difference between our experiment and Curran and Hintzman's experiments was that we required our participants to pronounce study words aloud, as we have usually done, to ensure that they attended to each of the study words. As did Curran and Hintzman, we intermixed inclusion and exclusion tests for both the direct-retrieval and generate–recognize conditions.

Buchner, Erdfelder, & Vaterrodt-Plünnecke (1995) proposed a multinomial model that allows base-rate differences to be taken into account when using the process-dissociation procedure. Their model treats guessing as independent of automatic influences of memory and uses performance on new items to estimate the probability of correct responding on the basis of guessing. Yonelinas and Jacoby (1996) compared results produced by different models of response bias used to deal with base rates. If base rates do not differ across conditions and one's only interest is dissociations, then a choice among models of response bias is not necessary. Using a multinomial model (Buchner et al., 1995) to separate base rate from automatic influences of memory does not change the pattern of results but does produce an estimate of automatic influences of memory that differs in magnitude from that gained by subtracting base rate from A. However, if base rate is lower for the exclusion test than for the inclusion test because of criterion differences, R will be inflated by an amount that reflects the difference in base rate between the two types of test, and A will be invalid because of the reliance of its estimation on R. To be successful, a model of response bias must "correct" for base-rate differences between inclusion–exclusion tests by producing results that are the same as observed when base rates do not differ across types of test.

For experiments reported here, we compared results gained by using our original estimation procedure with results gained by using the multinomial approach proposed by Buchner et al. (1995). Reliance on a generate–recognize strategy as opposed to a direct-retrieval strategy changes the nature of task performance in a way that invalidates our estimation procedure as well as creating base-rate differences. Consequently, we did not expect the paradoxical dissociation in the generate–recognize condition to be eliminated when the multinomial approach was used to correct for base-rate differences. As we show, the utility of a multinomial approach is as reliant on assumptions as is the original estimation procedure, and its use to correct for base-rate differences is not legitimate if assumptions underlying the process-dissociation procedure are violated.

Method

Participants. Forty-eight students participated in the experiment in return for credit in an introductory psychology course at McMaster University. Half of the participants were randomly assigned to the direct-retrieval test condition, and the other participants were assigned to the generate–recognize test condition. Nine additional participants were tested, but their data were discarded for purposes of analyses. Two of these participants were in the direct-retrieval condition. One of those participants had perfect performance (1.0) in the inclusion-long study condition, and the other participant had perfect performance (zero) in the exclusion-long study condition. The remaining 7 participants, whose data were discarded, were in the generate–recognize condition. Three of those participants had a zero in the exclusion-long study condition, 1 participant had a 1.0 in the inclusion-long study condition, and the remaining 3 participants failed to understand or to follow instructions. Participants in both conditions were asked at the end of the experiment whether they had given old words when cued with the prompt *new?* Three participants in the generate–recognize condition answered that question in a way that was interpreted as their failing to follow instructions. Two participants reported that they completed the stems with the first word that came to mind regardless of whether the prompt was *old* or *new.* The other participant reported responding with old words when he or she could not think of an alternative completion.

Materials and procedure. Materials were largely the same as those used in Experiment 1. Experiment 3 was actually done before Experiment 1, and materials were rebalanced (broken into different subsets) for Experiment 1 because the base rates for some of the completion words for the stems when new were slightly different for the Texas participants, as compared with the Ontario participants (e.g., *sma*—; "smash": Texas .28, Ontario .04; "small": Texas .07, Ontario .41). The construction of lists and procedure was the same as for the direct-retrieval condition in Experiment 1, except for the manipulation in the study phase. Rather than manipulating attention during study, presentation duration (10 s vs. 1 s) for study was manipulated as was done in Curran and Hintzman's (1995) experiments.

Two study lists of 50 words each (40 critical items and 10 buffer items) were presented with items in one list presented for 10 s each and items in the other list for 1 s each. The test list consisted of 120 three-letter word stems with 40 stems representing each of the types of word (long study, short study, and new). For each word type, half of the stems were presented in the inclusion-test condition, and half were presented in the exclusion-test condition. The inclusion- and exclusion-test items were intermixed and cued by the prompts *old* and *new* as in the direct-retrieval test conditions in Experiment 1.

Instructions for the study phases were the same as for Experiment 2, and instructions for the direct-retrieval test condition were similar to those in Experiment 1. Instructions for the generate–recognize test condition were also similar to those in Experiment 1 but were modified to accommodate the intermixing of inclusion and exclusion tests. In the generate–recognize test condition, participants were told that if a stem appeared with the cue word *old,* it was all right to complete the stem with a previously studied word, but it was not necessary to do so. They were instructed to complete those stems as quickly as possible with the first word that came to mind. In contrast, if the cue word was *new,* participants were told not to use an old word but rather to complete the stem with a new word. They were instructed to check their memory to be sure that a completion that came to mind was not one that had been earlier studied. If the completion word seemed at all familiar, they were not to use it but, instead, were to try to think of a different completion. If they were unable to think of a different completion, they were to leave the stem incomplete and let the 15-s completion deadline elapse.

Results and Discussion

Proportion of stems completed with old words. Performance on the inclusion test (Table 1) was higher for the 10-s than for the 1-s study condition, $F(1, 46) = 8.39$, $MSE = 0.014$. Although the interaction of instruction and study duration was not significant, the advantage of longer study was numerically larger in the direct-retrieval test condition than in the generate–recognize test condition. For the exclusion test, participants were less likely to mistakenly use an old word as a completion in the generate–recognize test condition, as compared with direct retrieval, $F(1, 46) = 15.29$, $MSE = 0.026$. Exclusion performance was also more accurate for the 10-s study condition than for the 1-s, $F(1, 46) = 35.96$, $MSE = 0.010$. In the direct-retrieval condition, the difference in baseline completion rates for the inclusion and exclusion tests did not approach significance, $F < 1$. However, in the generate–recognize condition, base rate was much lower for the exclusion test than for the inclusion test, $F(1, 23) = 8.93$, $MSE = 0.012$. The overall pattern of results shows the expected differences between direct-retrieval and generate–recognize strategies.

Estimates of R and A. The probability of recollection (Table 2) was higher in the generate–recognize test condition than in direct retrieval, $F(1, 46) = 4.36$, $MSE = 0.057$. For both test conditions, recollection was higher for the 10-s study condition than for the 1-s, $F(1, 46) = 34.07$, $MSE = 0.027$.

The analysis of A revealed a significant interaction between study duration and test conditions, $F(1, 46) = 5.23$, $MSE = 0.018$. The results of the direct-retrieval condition showed a process dissociation that is of the same form as found in Experiments 1 and 2. Shortening the study duration from 10 s to 1 s decreased R, $F(1, 23) = 23.07$, $MSE = 0.024$, but left A almost perfectly invariant, $F < 1$. This result did not reflect insensitivity of our measure because the power to detect an effect on A as large as the effect on A observed in the generate–recognize condition was .90 (Cohen's $d = 0.88$). Estimates of A for old items were significantly above base rate, $F(1, 23) = 53.28$, $MSE = 0.007$, showing a large effect of study on automatic influences of memory.

A paradoxical dissociation was found in the generate–recognize condition. Reducing study time from 10 s to 1 s decreased R, $F(1, 23) = 12.53$, $MSE = 0.031$, but had the opposite effect of increasing A, $F(1, 23) = 6.90$, $MSE = 0.021$. For words studied at the 1-s duration, A was significantly above baseline, $F(1, 23) = 8.26$, $MSE = 0.009$, but for the 10-s duration, A did not differ from baseline, $F <$

1. Participants' reliance on a generate–recognize strategy produced a paradoxical dissociation by violating both the independence assumption and the assumption of equality of R for inclusion and exclusion tests.

Multinomial model analyses. For each instructional condition, we gained parameter estimates by using a direct-retrieval multinomial model adapted from Buchner et al., (1995, see Appendix C). This model includes consciously controlled and automatic processes, as well as a third process, guessing, all of which are independent. Thus, we used six parameters, R_{long}, R_{short}, A_{long}, A_{short}, G_{inc}, and G_{exc} to fit frequency data from the six cells of each instructional condition (inclusion short, inclusion long, inclusion new, exclusion short, exclusion long, and exclusion new). There were a total of 2,880 observations (24 participants × 120 observations/participant) in each instructional condition. Separate fits were obtained for the direct-retrieval and generate–recognize instructional conditions.

The G-power program made available by Erdfelder, Faul, and Buchner (1996) was used to conduct a post hoc power analysis to obtain values of β and the critical value of χ^2. This analysis required input of four parameters, w (the "effect size" for chi-square tests; Cohen, 1977), N, α, and the degrees of freedom. We used $w = 0.10$, $N = 2,880$, $\alpha = 0.005$, and one degree of freedom. The power analysis yielded $\beta = 0.0052$ (power = .9948) and $\chi^2_{crit}(1) = 7.88$.

To test the independence assumption, we tested the fit of a restricted version of the model where $A_{long} = A_{short}$ (see Riefer & Batchelder, 1988, for a discussion of placing restrictions on parameters to reflect assumptions in a multinomial model). The fits of the model to the data were assessed with the maximum-likelihood statistic, G^2, computed by using the multinomial binary tree (MBT) program (Hu, 1995) and compared against a chi-square distribution with one degree of freedom. If the independence assumption of the multinomial model is valid and study duration does not influence A, then the restricted version of the model should fit the data from the direct-retrieval condition. In contrast, we expected that the model would not fit the data from the generate–recognize condition because the instructions given to participants in that condition induced process dependence, and therefore A_{long} should not equal A_{short}. A dependence version of the model would not have the restriction that $A_{long} = A_{short}$ and, of course, would fit the data, because without that restriction there are as many parameters as cells so that there are no degrees of freedom left for a test.

The fit of the restricted model to data from the direct-retrieval condition was extremely good, $G^2(1) = 0.52$. As with the ANOVA carried out on these data in the previous section, we failed to reject the hypothesis that $A_{long} = A_{short}$. In contrast, for the generate–recognize data, the fit of the model was poor, $G^2(1) = 9.16$, $p < .005$. The poor fit of the restricted model to the generate–recognize data necessitates rejection of the hypothesis that $A_{long} = A_{short}$. Again, the results of the multinomial analysis for the generate–recognize data parallel those of the ANOVA, where estimates of A_{long} were found to be significantly lower than estimates of A_{short}.

Multinomial analyses were also done by using the data from Experiment 1. Using the same direct-retrieval model (substituting the parameters R_{full}, R_{div}, A_{full}, A_{div}, G_{inc}, and G_{exc}), we fit the data from the six cells of each instructional condition in Experiment 1 (inclusion–divided, inclusion–full, inclusion–new, exclusion–divided, exclusion–full, exclusion–new).[2] As before, we tested the independence assumption of the model by using a restricted version of the model where $A_{full} = A_{div}$. Results of these analyses paralleled results of the ANOVA on estimates of A in Experiment 1; the model fit the data from the direct-retrieval condition exceptionally well, $G^2(1) = 0.43$, whereas the fit of that same direct-retrieval model to the generate–recognize condition was poor, $G^2(1) = 21.61$, $p < .005$. Parameter estimates generated by the direct-retrieval model are shown in Appendix D.

Can we find a model that will fit the data from the generate–recognize condition? To do so, we rearranged the parameters in the direct-retrieval model to more closely reflect operation of a generate–recognize strategy (see Appendix C). As with the direct-retrieval model, we attempted to fit the generate–recognize model to data from both instructional conditions in Experiments 1 and 3. The generate–recognize model (Appendix C) contained the same parameters as the direct-retrieval model: R_{full}, R_{div}, A_{full}, A_{div}, G_{inc}, and G_{exc} in Experiment 1 and R_{long}, R_{short}, A_{long}, A_{short}, G_{inc}, and G_{exc} for Experiment 3. For both experiments, $\alpha = .005$, and $\chi^2_{crit}(1) = 7.88$. For Experiment 1, analysis of the restricted model where $A_{full} = A_{div}$ indicated a good fit of the model for the generate–recognize condition, $G^2(1) = 3.33$, but not for the direct-retrieval condition, $G^2(1) = 13.78$, $p < .005$. Likewise, for Experiment 3, the restricted model where $A_{long} = A_{short}$ fit the data for the generate–recognize condition quite well, $G^2(1) = 1.41$, but did not fit the data for the direct-retrieval condition, $G^2(1) = 13.43$, $p < .005$. Parameter estimates generated by the generate–recognize model are shown in Appendix D.

Although participants sometimes use a generate–recognize strategy to accomplish cued recall, we have little interest in developing a generate–recognize model. The assumptions underlying a generate–recognize model are no less open to violation than are those underlying a direct-retrieval model (e.g., Weldon & Colston, 1995). A generate–recognize model (Jacoby & Hollingshead, 1990) rests on an independence assumption, although one that is different from the direct-retrieval model. Also, we are hesitant to adopt a generate–recognize model because doing so relies on the assumption that an inclusion test (indirect test) provides a process-pure measure of automatic processes (see the exchange between Jacoby, Toth, Yonelinas, & Debner,

[2]There were 5,760 observations (48 participants × 120 observations/participant) in each instructional condition in Experiment 1. As in Experiment 3, separate fits were obtained for the direct-retrieval and generate–recognize instructional conditions. A post hoc power analysis was conducted to obtain values of β and the critical value of χ^2. With $w = 0.10$, $N = 5,760$, $\alpha = 0.005$, and one degree of freedom, the power analysis yielded $\beta = 0.005$ (power = .995) and $\chi^2_{crit}(1) = 7.88$.

1994, and Joordens & Merikle, 1993). Most important, our primary interest is in recollection. To satisfy the assumptions of a generate–recognize model, we must attempt to ensure that participants do not use recollection to accomplish cued recall. Instructing participants to engage in recollection for the inclusion test, as done for the direct-retrieval condition, changes the nature of task performance in a way that violates assumptions underlying the generate–recognize model.

Clearly, base-rate differences between the inclusion and exclusion tests cannot be treated merely as reflecting differences in guessing and then "corrected" for guessing by adding an independent guessing parameter to the model. Rather, the retrieval strategy used by participants determines how "guessing," and other processes, operate within the framework of the model. Adopting a generate–recognize strategy violates assumptions underlying the process-dissociation procedure, and results cannot be rectified by correcting for guessing.

Correlations. No correlation at the level of participants was significant, and only in the generate–recognize condition was the correlation at the level of items for $R_{old} - A_{old}$ significant (.28). The reason that correlations were lower in this experiment than in Experiment 1 may be a result of the fact that a smaller number of participants was tested, and so estimates were based on fewer observations. At both the participant level and the item level, estimates of A were near identical for long and short study durations (.59 vs. .58, when based on participant means, and .53 vs. .51, when based on item means). The small correlations observed between R and A were similar in magnitude and opposite in direction for participants and items, yet estimates gained in the two ways showed the same pattern of results.

Comparisons with results reported by Curran and Hintzman (1995). Only in Curran and Hintzman's Experiment 5 did a paradoxical dissociation remain when the inclusion–exclusion procedure was used and zero scores were dropped. The pattern of results from that experiment is strikingly similar to results found for the generate–recognize condition (see Tables 1 and 2). The significant difference in base rate between the inclusion and exclusion tests found in Curran and Hintzman's Experiment 5 suggests that their participants relied on a generate–recognize strategy just as did our participants who were instructed to do so. The magnitude of the paradoxical dissociation observed in Curran and Hintzman's Experiment 5 is nearly identical to that produced by our generate–recognize instructions.

Results in the direct-retrieval condition were similar to those from Curran and Hintzman's (1995) Experiment 4, showing that increasing study time increased R but left A relatively invariant. Curran and Hintzman (1995) argued that the absence of a paradoxical dissociation in their Experiment 4 was because participants did not understand exclusion instructions (but see the exchange between Jacoby, Begg, & Toth, 1997, and Curran & Hintzman, 1997, for debate concerning the criteria that they used to conclude that participants did not understand instructions). There is no evidence to suggest that participants in our experiment did not understand exclusion instructions.

General Discussion

The three experiments produced extremely consistent results. Direct-retrieval instructions produced process dissociations, as did direct-retrieval data with the IRK procedure. Decreasing study time reduced recollection but left automatic influences almost perfectly invariant (Experiment 2 and 3), just as did dividing attention during study (Experiment 1; Jacoby et al., 1993; Schmitter-Edgecombe, 1996). When generate–recognize instructions were given or generate-recognize performance was mimicked with the IRK procedure, a paradoxical dissociation was produced by manipulating attention (Experiment 1) and by manipulating study time (Experiments 2 and 3)—increases in R were accompanied by a decrease in A. Paradoxical dissociations result when participants use a generate–recognize strategy, thereby excluding items that they "know" are old. To meet the assumptions underlying the process-dissociation procedure, it is important that instructions encourage participants to exclude items only because they recollect (remember) earlier studying the items.

Effects of Instructions

Because of automatic influences of memory, reading a word makes it more likely that the word will later come readily to mind as a completion for a word stem. Clearly, excluding old words because of the fluency with which they came to mind as a completion would violate the independence assumption underlying the estimation procedure. Fluency reflects automatic influences of memory and so, as a basis for exclusion, cannot be independent of those influences. Elsewhere, we (e.g., Jacoby, Kelley, & Dywan, 1989) have argued that the feeling of familiarity that can serve as an alternative to recollection as a basis for recognition memory reflects an unconscious attribution process that relies on a fluency heuristic. Applying those arguments, when a generate–recognize strategy is used or when asked to make remember–know judgments, we can correctly attribute fluency to its source and experience it as familiarity or "knowing," which can be used to exclude old words. However, the attribution process is influenced by instructions. When a direct-retrieval strategy is used, the same automatic influences of memory on fluency can be ignored or attributed to differences among items and not experienced as familiarity. Fluency of completing a stem is ambiguous in that it does not specify its source. Because of its ambiguity, reliance on fluency, as a basis for recognition or knowing that is used to exclude old words, results in the exclusion of new words that are fluently produced as completions as well as old words, creating a difference in base rate between inclusion and exclusion tests.

Richardson-Klavehn and Gardiner (1996) have suggested that because of involuntary conscious memory, the process-dissociation procedure confounds awareness with intention and underestimates automatic influences. Their argument is that old words are sometimes excluded, although they are not intentionally brought to mind by means of recollection but, rather, involuntarily come to mind as a completion followed by awareness of their prior study. Important for their argument is the question of whether awareness of prior study (recognition memory) is involuntary. If it is not, exclusion because of involuntary conscious memory does not differ from the use of a generate–recognize strategy to exclude old words and depends on instructions that are used. Involuntary recognition memory does sometimes occur (Ste-Marie & Jacoby, 1993), but it is yet to be shown that it plays a role in the inclusion–exclusion procedure (Reingold & Toth, 1996). Involuntary conscious memory of an item may seldom occur in the context of intentionally trying to remember the item. Also, estimates would not be influenced by any involuntary conscious memory that occurs after an old word is given as a completion or, for other reasons, is not used as a basis for exclusion.

Combining the direct-retrieval and generate–recognize strategies would allow participants to be more certain of excluding old words than would the use of either strategy alone. Using this combined strategy, participants would attempt to recall an old word by using the stem as a cue and, if unsuccessful, would then think of a completion that was subjected to a voluntary, recognition-memory check before being output as a response. Our direct-retrieval instructions do not tell participants to use a recognition-memory check. Rather, participants are told only to exclude words that they recall as earlier presented. Convergence of results from the inclusion–exclusion procedure with those from the IRK procedure supports the claim that use of a recognition-memory check is voluntary and that participants given direct-retrieval instructions are following those instructions. Participants given direct-retrieval instructions do not exclude words that they would "know" were old had they made remember–know judgments.

The problems for estimating automatic influences of memory said to be created by involuntary conscious memory are the same as those produced by participants' use of a generate–recognize strategy to exclude old words. The goals of direct-retrieval instructions are to satisfy the assumption that R is the same for inclusion and exclusion tests as well as to satisfy the independence assumption. We suspect that intermixing the two types of test makes it more likely that assumptions will be satisfied (cf. Buchner et al., 1995). We recommend that those who want to replicate our results use our direct-retrieval instructions (Appendix B) as a guide, keeping in mind that participants' correct understanding of instructions will result in their excluding only items that are recollected as earlier studied.

Remember–Know and Process Dissociation

The paradoxical dissociation produced by grouping "know" responses with "remember" responses and by using the IRK procedure is the same as found when using generate–recognize instructions with the inclusion–exclusion procedure and is also similar to the paradoxical dissociation that Curran and Hintzman (1995) found by using their recollect and exclude procedure. In all these cases, only old words that were misclassified as "new" would be used as completions on the exclusion test. The high accuracy of exclusion performance reflects exclusion on the basis of recognition reflecting familiarity ("know" responses), rather than exclusion being restricted to the use of recollection ("remember" responses), as is required to meet assumptions underlying the process-dissociation procedure and for correct application of the IRK procedure. Misclassifying as "remember" words that participants only "know" are old inflates estimates of R and produces a paradoxical dissociation by producing decreases in A with increases in R.

Whereas estimation of automatic influences by using our IRK procedure is based on an independence assumption, Gardiner and his colleagues (e.g., Gardiner, 1988; Gardiner & Java, 1991; Gardiner & Parkin, 1990) have analyzed the straight probability of a "know" response. The choice between approaches is important for the finding of dissociations (Jacoby, Yonelinas, & Jennings, 1997; Yonelinas & Jacoby, 1996). The remember–know procedure has most often been used in conjunction with tests of recognition memory, whereas for the experiments reported here, we used the procedure in combination with cued recall. Recognition memory and cued recall differ with regard to the dissociations they reveal (Jacoby, Yonelinas, & Jennings, 1997). Also, our primary interest has been in separating contributions of automatic and consciously controlled processing, whereas Gardiner and his associates approach similar issues from a phenomenological standpoint (Richardson-Klavehn & Gardiner, 1996).

A cost of disagreement is that it sometimes overshadows agreement. We agree with Gardiner and his colleagues regarding the importance of subjective reports and the utility of the remember–know procedure. It should be noted that the correspondence between the probability of "remember" and R as measured by inclusion–exclusion remains regardless of the assumption made about the relation between consciously controlled and automatic processes. Both the remember–know and the process-dissociation approach reject a unitary view that does not distinguish different forms or uses of memory.

The remember–know approach has recently been criticized by showing that dissociations that it reveals can be accounted for by a single-process model of memory in combination with signal-detection theory (Donaldson, 1996; Hirshman & Master, 1997). Convergence between results from the remember–know and process-dissociation procedures can be helpful for showing the need to distinguish between uses or forms of memory. However, such convergence should not always be expected because the inclusion–exclusion procedure measures R as that which affords control over responses, and the IRK procedure measures phenomenological experience. Also, for both the remember–know and the process-dissociation procedures, details of the instructions that are used are important.

Base-Rate Differences as Direct Evidence of Violated Assumptions

The presence of a significant difference in base rates between inclusion and exclusion tests can be used as direct evidence that assumptions underlying the estimation procedure have been violated. Generate–recognize instructions produced significant differences in base rates, as did mimicking a generate–recognize strategy with the remember–know procedure. However, a problem for the inclusion–exclusion procedure is that the absence of significant base-rate differences between types of test does not ensure that assumptions underlying the estimation procedure have been met.

Curran and Hintzman (1995) found paradoxical disssociations that were not accompanied by base-rate differences. In their Experiment 1, the paradoxical dissociation was eliminated when zeros in exclusion were dropped, suggesting that the paradoxcial dissociation was caused by floor effects rather than by violation of asssumptions. However, for their Experiment 5, dropping zero scores in exclusion reduced the base-rate difference between inclusion and exclusion tests to the point it was no longer significant but left a paradoxical dissociation. Russo and Andrade (1995) also found paradoxical dissociations that were not accompanied by significant differences in base rates. In their experiments, materials were used that produced very low base rates and exclusion of old words was very near perfect. Even without zero scores, problems are produced for estimating A when performance is very near floor. At that extreme, estimates of A are much more sensitive to small differences in exclusion performance.

Paradoxical dissociations sometimes reflect floor effects rather than the violation of assumptions underlying the estimation procedure. The choice between means of dealing with zero scores in exclusion performance is controversial (see Jacoby, Begg, & Toth, 1997, and the response by Curran & Hintzman, 1997). However, it is important that using materials that produce a higher base rate so as to avoid zero scores eliminates "paradoxical" dissociations that were otherwise found (e.g., Jacoby et al., 1993). Also, the use of materials that produce higher base rates is likely to increase the sensitivity of base-rate differences to participants' use of a generate–recognize strategy. When base rates are very low, even if participants are using a generate–recognize strategy, few new words are likely to come fluently to mind as a completion and to be mistakenly excluded because of their false recognition, and such mistaken exclusion is necessary to produce significant base-rate differences. We recommend that investigators select materials (Appendix A) to produce base rates that are sufficiently high to avoid zeros in exclusion. However, even when higher base rates are used, participants' reliance on a generate–recognize strategy might not always produce a significant base-rate difference.

Process Dissociation and Multinomial Models: Correcting for Base-Rate Differences

Buchner et al. (1995) advanced an "extended measurement model" for the process-dissociation procedure that was said to have the advantage of taking base-rate differences in guessing into account. The results reported here provide reasons for caution when using such a model. The difficulty is that base-rate differences may not simply reflect differences in guessing but, rather, reflect a difference in ways the task is being accomplished. When generate–recognize instructions were given, use of a multinomial model did not correct for base-rate differences in a way that eliminated the paradoxical dissociation that was produced.

Using a multinomial model does have some potential advantages as a means of analyzing data. For example, it makes it clear that we are testing a model and helps rebut claims to the contrary (e.g., Hillstrom & Logan, 1997; Richardson-Klavehn et al., 1996). However, as a general class, multinomial models allow a large variety of different assumptions and are atheoretical, providing only a means of describing data (Batchelder & Riefer, 1990). By the process-dissociation approach, in contrast, we use theory to make an explicit assumption about the relation between processes, to design conditions, and to make predictions meant to test our assumptions. For example, process dissociations showing effects on R in combination with relative invariance of A should be consistently found only if R and A are independent.

We have developed instructions and experimental procedures that are meant to avoid base-rate differences between inclusion and exclusion tests. If base rates do not differ and one's interest is in dissociations, a model of response bias is unnecessary. A disadvantage of the multinomial approach is that investigators may be misled to believe that the multinomial model provides a statistical means of accomplishing equal base rates, making it unnecessary to avoid base-rate differences by design. Yonelinas and Jacoby (1996) provided a more complete discussion of the relation between multinomial models and the process-dissociation approach and argued for the advantages of avoiding base-rate differences.

Can Correlations Be Used as Direct Evidence of Violated Assumptions?

Curran and Hintzman (1995) interpreted findings of a positive correlation between R and A at the level of items as direct evidence that the independence assumption underlying the process-dissociation procedure was invalid. Because of the positive correlation between R and A, the underestimation of A was said to increase with increases in R and, thereby, produce a paradoxical dissociation. Curran and Hintzman (1997) stressed the importance of the correlation between R_{old} and A_{new} at the level of items as suggesting that correlation at the level of items reflects "common determinants of lexical access" (p. 501). Hintzman and Curran (1997, p. 512) stated that correlations across subjects and correlations across items are of no relevance to the unobservable correlation that measures process dependence but are "direct evidence" that the correlation responsible for aggregation bias is unlikely to be zero and serve as a warning that "the possibility of underestimation should not be ignored."

In our experiments, the correlation between R_{old} and A_{new} never approached significance in the direct-retrieval conditions, and so lexical access cannot be credited as the source of the correlation at the item level in those conditions. Rather, the correlation between R_{old} and A_{old} at the item level in direct-retrieval conditions might have been produced by the estimation procedure itself. Aggregating across participants to compute item-level means should produce overestimation of A, rather than underestimation, because of the inverse correlation at the participant level. If it is to be held that observed correlations between estimates serve as direct evidence that aggregation bias played a role in results found with direct-retrieval instructions, it must be explained why results were the same regardless of whether item-based or participant-based estimates were used, although effects of aggregation bias should be opposite for the two types of estimate. Curran and Hintzman's (1995) Experiment 4 also showed effects on R but relative invariance in A for both item-based and participant-based means, although the correlation between R and A was positive at the item level and inverse at the participant level.

Results are consistent with Jacoby and Shrout's (1997) conclusion that observed correlations at the item or participant level cannot be used to directly test process dependence or to infer the unobservable correlations responsible for aggregation bias. The observed correlations do not allow the prediction of whether a paradoxical dissociation will be found. When the experiments avoided zeros in exclusion and used direct-retrieval instructions or mimicked use of a direct-retrieval strategy with the remember–know procedure, there were no paradoxical dissociations left to be explained, but significant correlations were still found. Correlations at the item level were nearly as large in the direct-retrieval conditions as in the generate–recognize conditions, although a paradoxical dissociation was found only in the latter conditions. A significant correlation at the level of items between R_{old} and A_{new} was found only for generate–recognize results (Experiments 1 and 3). However, comparisons across our results and those of Curran and Hintzman (1995, Experiment 4) show that, with or without a significant correlation of R_{old} and A_{new} at the level of items, it is possible to find an increase in R along with relative invariance of A (i.e., an absence of a paradoxical dissociation).[3] A significant correlation between R and A at the level of participants was not found when generate–recognize instructions were used, providing empirical evidence that joins Jacoby and Shrout's (1997) psychometric analysis to show that the significance of that correlation cannot serve as a direct test of whether or not participants were using a generate–recognize strategy.

Hintzman and Curran (1997) said, "We think it is unlikely that subjects in process-dissociation experiments on word-stem completion can be prevented from using a generate–recognize strategy at least some of the time" (p. 513). Curran and Hintzman (1995) interpreted the inverse correlation between R and A at the level of participants as evidence that participants relied on a generate–recognize strategy. However, Hintzman and Curran agreed with Jacoby and Shrout (1997) that correlation at the participant or item level cannot be used to test process dependence such as that produced by participants' reliance on a generate–recognize strategy. If participants in our direct-retrieval conditions did sometimes use a generate–recognize strategy, their doing so was not sufficient to produce a paradoxical dissociation. To argue that the independence assumption was violated in a way that produced either process dependence or aggregation bias, one has to argue that, in some mysterious way, participants' misunderstanding of instructions or a true effect of the manipulated variable on A consistently served to offset near perfectly the effect of violation of independence, regardless of whether item-level or participant-level means were examined. We believe it is more likely that participants understood and followed instructions in the direct-retrieval conditions than that such delicate balances between offsetting effects can be routinely found.

Understanding of correlations between R and A, at whatever level, requires a thorough psychometric analysis along with identification of the many potential sources of correlation, including the contribution of the estimation procedure itself (Jacoby & Shrout, 1997). We are not tempted to try to interpret significant correlations that were found because of the potential contribution of the estimation procedure as well as that of other sources of correlation that have not yet been identified. Correlations found in the generate–recognize conditions are particularly difficult to interpret because of violations of assumptions underlying the process-dissociation procedure. However, if one is interested in individual differences, it is important to gain a better understanding of the sources of correlation that are responsible for the significant correlations observed in the direct-retrieval conditions. Those correlations cannot be used to directly test assumptions underlying the process-dissociation procedure but are sufficiently high to warrant their further investigation.

The Process-Dissociation Approach: Limits and Future Directions

Results from the direct-retrieval conditions coverge with a great deal of other evidence to suggest that it is possible to satisfy the independence assumption underlying the process-dissociation procedure (Jacoby, Yonelinas, & Jennings, 1997). Hay and Jacoby (1996) created conditions necessary to apply the process-dissociation procedure by varying the relation between material that was to be remembered and prior knowledge rather than by use of inclusion–exclusion instructions. Doing so avoids many potential problems, such as differences in base rate, and produces the same pattern of results as does the inclusion–exclusion procedure. Ratcliff and McKoon (1997) presented a countermodel that describes effects of implicit memory on perceptual performance (e.g., Jacoby & Dallas, 1981) as produced by bias. We (Jacoby, McElree, & Trainham, in press) have shown that results reported by Ratcliff and McKoon, when reanalyzed, reveal striking process dissociations that are the same as found by Hay and Jacoby (1996). The countermodel provides a description of a more complete information-processing model that is compatible with the process-

[3] It is unclear why a significant correlation between R_{old} and A_{new} was found in Curran and Hintzman's (1995) Experiment 4 but was not found in the direct-retrieval condition in our experiments. However, to prove that correlation provides a direct test of the independence assumption requires showing that the correlation could not have arisen from some other source, including aggregation, which we believe is impossible to accomplish.

dissociation approach and provides exciting new directions for research.

Results described above, along with those from experiments reported here, should be sufficient to discourage abandoning the process-dissociation approach. We continue to be impressed by results obtained by using the inclusion–exclusion procedure when the boundary conditions specified by Jacoby et al. (1993) are met. When zeros in exclusion are avoided and direct retrieval is encouraged, process dissociations showing effects on R accompanied by near perfect invariance in A are consistently found. However, on the negative side, details of instructions serve as an important boundary condition for findings, and there is no fully reliable means of directly testing assumptions underlying the estimation procedure. We see those problems as a challenge to further refine procedures and to explore boundary conditions rather than as reasons to abandon the goal of separating the contributions of automatic and controlled processes.

Even if the assumptions of the process-dissociation procedure hold across only a very limited range of conditions, the procedure might still be useful for diagnosis and the design of treatments for deficits in memory (Jacoby, Jennings, & Hay, 1996). It is deficits in recollection that we find most interesting and seek to measure by arranging situations such that participants rely on a direct-retrieval strategy rather than on a generate–recognize strategy. The assumption made about the relation between automatic and controlled processes is important for attempts to remediate memory deficits. Suppose that because of a severe memory deficit, a person was unable to correctly answer a question about what was eaten for breakfast. A direct-retrieval approach to rehabilitation would encourage an attempt to train recollection. For example, the retention interval might be varied and the person trained to better constrain retrieval by attempting to reconstruct context. In contrast, a generate–recognize approach would encourage the person suffering a memory deficit to let an answer automatically come to mind and then carefully inspect the potential answer to be absolutely certain that is was recognized as correct before giving it as a response. The generate–recognize approach might sometimes be effective but seems unattractive. As those who have had critics know, inspecting one's answer to be absolutely certain that it is correct before giving it as a response can produce overvigilance that is extremely unpleasant and can be counterproductive. Fof memory performance, as well as for theorizing about memory, mistaken exclusion produced by overvigilance can be far more problematic than the failures to exclude that the greater vigilance was meant to eliminate.

References

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97,* 548–564.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124,* 137–160.

Challis, B. H., and Brodbeck, D. R. (1992). Level of processing affects priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 595–607.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 4,* 335–359.

Cohen, J. (1977). *Statistical analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Craik, F. I. M. (1982). Selective changes in encoding as a function of reduced processing capacity. In F. Klix, J. Hoffman, & E. van der Meer (Eds.), *Cognitive research in psychology* (pp. 152–161). Berlin, Germany: Deutscher Verlag der Wissenschaffen.

Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 531–547.

Curran, T., & Hintzman, D. L. (1997). Consequences and causes of correlations in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 496–504.

De Houwer, J. (in press). Differences in conscious retrieval during inclusion and exclusion tasks. *Memory.*

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 9,* 1–19.

Erdfelder, E., Faul, F., & Buchner, A. (1996). G-Power: A general power analysis program. *Behavior Research Methods, Instruments and Computers, 28,* 1–11.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition, 19,* 617–623.

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18,* 579–583.

Graf, P., & Komatsu, S. (1994). Process dissociation procedure: Handle with caution! *European Journal of Cognitive Psychology, 6,* 113–129.

Greene, R. L. (1986). Word stems as cues in recall and completion tasks. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38*(A), 663–673.

Hay, J. F., & Jacoby, L. L. (1996). Separating habit and recollection: Memory slips, process dissociations, and probability matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1323–1335.

Hillstrom, A. P., & Logan, G. D. (1997). Process dissocation, cognitive architecture, and response time: Comments on Lindsay and Jacoby (1994). *Journal of Experimental Psychology: Human Perception and Performance, 20,* 219–234.

Hintzman, D. L., & Curran, T. (1997). More than one way to violate independence: Reply to Jacoby and Shrout (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 511–513.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory. Reflections on the remember–know paradigm. *Memory & Cognition, 25,* 345–351.

Hu, X. (1995). *Statistical inference program for multinomial binary tree models.* [Computer program]. Irvine, CA: University of California, Irvine.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika, 59,* 21–48.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

Jacoby, L. L., Begg, I. M., & Toth, J. P. (1997). In defense of functional independence: Violations of assumptions underlying the process-dissociation procedure? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 484–495.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3,* 306–340.

Jacoby, L. L., & Hay, J. F. (in press). Age-related deficits in memory: Theory and application. In M. A. Conway (Ed.), *Theories of memory II.* Mahwah, NJ: Erlbaum.

Jacoby, L. L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language, 29,* 433–454.

Jacoby, L. L., Jennings, J. M., & Hay, L. F. (1996). Dissociating automatic and consciously controlled processes: Implications for diagnosis and rehabilitation of memory deficits. In D. J. Herrmann, C. L. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research: Theory in context* (Vol. 1, pp. 161–193). Mahwah, NJ: Erlbaum.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., McElree, B., & Trainham, T. N. (in press). Automatic influences as accessibility bias in memory and Stroop-like tasks: Towards a formal model. In A. Koriat & D. Gopher (Eds.), *Attention and performance XVII.*

Jacoby, L. L., & Shrout, P. E. (1997). Toward a psychometric analysis of violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 505–510.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122,* 139–154.

Jacoby, L. L., Toth, J. P., Yonelinas, A. P., & Debner, J. A. (1994). The relationship between conscious and unconscious influences: Independence or redundancy? *Journal of Experimental Psychology: General, 123,* 216–219.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. (1997). The relation between conscious and unconscious (automatic) influences. A declaration of independence. In J. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Mahwah, NJ: Erlbaum.

Joordens, S., & Merikle, P. M. (1993). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General, 122,* 462–467.

Koriat, A., & Feuerstein, N. (1976). The recovery of incidentally acquired information. *Acta Psychologica, 40,* 463–474.

Lindsay, D. S., & Kelley, C. M. (1996). Creating illusions of knowing in a cued recall remember/know paradigm. *Journal of Memory and Language, 35,* 197–211.

Mäntylä, T. (1993). Knowing but not remembering: Adult age differences in recollective experience. *Memory & Cognition, 21,* 379–388.

Parkin, A. J., Reid, T. K., & Russo, R. (1990). On the differential nature of implicit and explicit memory. *Memory & Cognition, 18,* 507–514.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104,* 319–343.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339.

Reingold, E. M., & Toth, J. P. (1996). Process dissociations versus task dissociations: A controversy in progress. In G. Underwood (Ed.), *Implicit cognition* (pp. 159–202). Oxford, England: Oxford University Press.

Richardson-Klavehn, A., & Gardiner, J. M. (1996). Cross-modality priming in stem completion reflects conscious memory, but not voluntary memory. *Psychonomic Bulletin & Review, 3,* 238–244.

Richardson-Klavehn, A., Gardiner, J. M., & Java, R. I. (1996). Memory: Task dissociations, process dissociations and dissociations of consciousness. In G. Underwood (Ed.), *Implicit cognition* (pp. 85–158). Oxford, England: Oxford University Press.

Roediger, H. L., III, & McDermott, K. B. (1993). Implicit memory in normal human subjects. In H. Spinnler & F. Boller (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63–131). Amsterdam: Elsevier.

Russo, R., & Andrade, J. (1995). The directed forgetting effect in word-fragment completion: An application of the process dissociation procedure. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 405–423.

Schmitter-Edgecombe, M. (1996, November). *Effects of divided attention on perceptual and conceptual memory test performances: A process dissociation framework.* Paper presented at the 37th Annual Meeting of the Psychonomic Society, Chicago, IL.

Schneider, W. (1990). *Micro-Experimental Laboratory.* [Computer program]. Pittsburgh, PA: Psychology Software Tools, Inc.

Ste-Marie, D. M., & Jacoby, L. L. (1993). Spontaneous versus directed recognition: The relativity of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 777–788.

Thorndike, E. L., & Lorge, I. (1944). *The teachers' handbook of 30,000 words.* New York: Columbia University Press.

Toth, J. P., Reingold, E. M., & Jacoby, L. L. (1994). Toward a redefinition of implicit memory: Process dissociations following elaborative processing and self-generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 290–303.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385–398.

Weldon, M. S., & Colston, H. L. (1995). Dissociating the generation state in implicit and explicit memory tests: Incidental production can differ from strategic access. *Psychonomic Bulletin & Review, 2,* 381–386.

Yonelinas, A. P., & Jacoby, L. L. (1996). Response bias and the process dissociation procedure. *Journal of Experimental Psychology: General, 125,* 422–439.

Appendix A

Multiple Completion Stems

Stem	Word completion	Base rate ON	Base rate TX	Word frequency	Set size[a]
act– –	actor	.70	.85	28	2
ali– –	alien	.31	.22	13	5
ang– –	angle	.44	.16	30	5
arg– –	argue	.82	.85	33	2
arr– –	arrow	.25	.32	37	2
bat– –	batch	.25	.35	3	4
ber– –	berry	.35	.32	29	5
bir– –	birch	.50	.31	16	2
bla– –	black	.31	.25	AA	8
bli– –	blind	.40	.38	A	5
blo– –	block	.33	.28	A	6
boa– –	boast	.47	.32	37	2
bre– –	bread	.51	.31	A	3
bri– –	bride	.37	.25	41	8
bro– –	broke	.42	.35	A	8
bud– –	buddy	.63	.82	3	2
bul– –	bully	.55	.38	10	3
cab– –	cabin	.55	.44	A	3
cam– –	camel	.80	.28	4	2
cau– –	cause	.79	.75	AA	2
cho– –	choke	.51	.38	27	5
chu– –	chunk	.75	.47	3	6
cla– –	clamp	.33	.35	4	8
cle– –	clerk	.50	.25	A	5
cli– –	click	.32	.63	10	5
clo– –	cloth	.36	.25	A	8
clu– –	clump	.33	.38	9	4
cou– –	couch	.25	.60	28	7
cra– –	crack	.41	.53	48	9
cri– –	crime	.26	.31	A	5
dan– –	dance	.61	.75	AA	2
del– –	delay	.34	.19	A	3
den– –	dense	.49	.41	19	2
dit– –	ditch	.58	.41	28	3
div– –	diver	.62	.69	4	5
dou– –	doubt	.43	.53	AA	3
dre– –	dream	.37	.44	AA	4
eag– –	eagle	.65	.50	38	2
ear– –	earth	.40	.44	AA	2
emb– –	ember	.24	.41	5	3
eth– –	ether	.47	.32	4	3
eve– –	event	.38	.22	A	2
fai– –	faith	.35	.38	A	3
fau– –	fault	.63	.63	A	2
fli– –	flick	.27	.35	4	4
for– –	forge	.31	.19	17	6
fre– –	freak	.29	.41	5	4
fro– –	frost	.33	.38	41	7
gla– –	glaze	.43	.25	9	5
gli– –	glide	.84	.57	41	3
glo– –	glory	.39	.19	A	6
goo– –	goose	.37	.60	45	3
gri– –	grind	.33	.44	18	6
gui– –	guide	.49	.41	AA	5
gul– –	gully	.37	.54	3	2
hat– –	hatch	.62	.51	19	3
hav– –	haven	.62	.61	3	2
hea– –	heavy	.22	.38	AA	5
hon– –	honey	.48	.29	A	3
hor– –	horse	.48	.51	AA	3
hou– –	house	.78	.88	AA	2
hum– –	human	.22	.16	AA	4
inl– –	inlet	.31	.47	5	3
kna– –	knack	.55	.66	4	2
kno– –	knock	.71	.79	A	4
lab– –	label	.53	.41	7	3
lap– –	lapel	.24	.22	1	2
lev– –	level	.67	.44	A	3
lim– –	limit	.45	.47	A	4
lun– –	lunch	.52	.53	39	3
mar– –	march	.41	.25	AA	3
mer– –	merry	.35	.35	38	4
mon– –	money	.66	.75	AA	3
mou– –	mouse	.50	.60	34	6
mov– –	movie	.40	.44	29	3
mus– –	music	.44	.47	AA	3
pan– –	panic	.50	.69	19	5
pat– –	patch	.34	.44	34	5
per– –	perch	.31	.19	23	3
pla– –	plate	.27	.25	A	7
por– –	porch	.57	.66	A	4
qua– –	quack	.43	.32	7	5
rab– –	rabbi	.23	.22	1	2
rad– –	radio	.61	.50	41	4
ran– –	ranch	.42	.28	20	3
reb– –	rebel	.88	.85	25	4
riv– –	river	.62	.60	AA	3
roa– –	roach	.46	.47	1	2
rou– –	route	.32	.44	A	5
sau– –	sauce	.80	.78	27	4
ser– –	serve	.43	.44	AA	2
shi– –	shift	.38	.19	A	6
sku– –	skunk	.50	.47	13	3
slu– –	slump	.35	.16	6	4
sma– –	small	.41	.07	AA	4
sna– –	snack	.28	.66	1	5
spo– –	spoke	.29	.25	AA	7
squ– –	squaw	.20	.07	3	5
sta– –	stand	.25	.25	AA	9
sti– –	stick	.53	.50	A	6
stu– –	study	.40	.32	AA	8
swa– –	swamp	.50	.38	29	9
swi– –	swing	.30	.19	A	9
tal– –	tally	.32	.38	2	4
thi– –	thick	.40	.57	A	7
tor– –	torch	.25	.60	17	6
tra– –	tramp	.38	.03	24	9
tre– –	treat	.60	.50	A	5
tro– –	troop	.28	.10	A	5
tru– –	truth	.75	.47	AA	7
twi– –	twist	.47	.66	42	5
val– –	value	.62	.66	AA	5
vig– –	vigor	.65	.81	19	2
wag– –	wagon	.66	.29	A	3
wat– –	water	.80	.63	AA	2
whe– –	wheat	.46	.35	A	4
wit– –	witty	.29	.22	24	2
wri– –	write	.39	.60	AA	3
yea– –	yeast	.59	.44	7	2
you– –	youth	.67	.57	AA	2

Note. Base rates for ON = Ontario, TX = Texas. There were only 32 observations in the Texas base rates, whereas the base rates for Ontario have been accumulated over many experiments. Word frequency is based on Thorndike & Lorge (1944); A and AA are high-frequency ratings with a median of 47.5 per million.

[a]Set size = the number of five-letter word completions (no plurals or proper names) for the stems that have actually been given by participants (rather than dictionary set size).

(*Appendixes continue*)

Appendix B

Test Instructions for Experiments 1 and 2

Experiment 1: Direct-Retrieval Instructions for Inclusion–Exclusion Intermixed

In this part of the experiment, you will be presented with word stems, that is, the first three letters of a 5-letter word. You are to use the word stems as cues for recall of words that were earlier presented in either of the lists you just read. For example, the word *slime* was in one of the lists you just read. In the test phase, the word stem *sli*– – would be presented as a cue for recall of *slime*. However, not all the stems that will be presented can be completed with an earlier-presented word and, so, recall of an earlier-presented word will not always be possible. If you are unable to recall a word, you are to complete the stem with the first 5-letter word that comes to mind that fits the stem. No plurals or proper names are allowed as completions.

This next part of the instructions is specific to cases in which you are able to recall an earlier-presented word that can be used as a completion for a stem. For example, suppose that given the stem *sli*– – you were able to recall that the word *slime* was presented earlier. Whether you are to use that recalled word as a completion depends on the type of test item you are given. Some of the stems will be presented with the prompt *old*. For those test items, you are to use your recalled word as a completion. Other stems will be presented with the prompt *new*. For those test items, you are not to use your recalled word as a completion. Rather, you are to complete those stems with a word different from your recalled word. Given *new sli*– –, you would not say "slime" but, rather, would give some other word such as "slice" as a completion for the stem. It will sometimes happen that you can recall an earlier-presented word but cannot think of a different completion for the stem. In those cases, simply leave the stem blank.

In summary, you will be given stems to use as cues for recall of words that were presented earlier. If you are able to recall a previously presented word, you are to use that recalled word as a completion if the stem is accompanied by the prompt *old*. However, if the stem is accompanied by the prompt *new*, you are not to use a recalled word as a completion. It will sometimes happen that you are unable to recall an earlier-presented word that would complete the stem. This is certain to be the case because some of the stems can only be completed with words that were not earlier studied. When you are unable to recall an earlier-presented word, complete the stem with the first word that comes to mind regardless of whether the stem is accompanied by the prompt *old* or the prompt *new*. Try to complete as many stems as possible. However, remember it is important that you try to use recalled words to complete stems accompanied by *old* and to not use recalled words to complete stems accompanied by *new*. You will have 15 s to complete each stem. If you have not completed the stem within 15 s, the program will simply go on to the next trial. Just give your response out loud.

Experiment 1: Generate–Recognize Instructions for Inclusion–Exclusion Blocked

Exclusion Test Instructions

The next part of the experiment involves a word-stem completion task. You will be presented with the first three letters of a 5-letter word, and your task is to complete the word. Some of the stems can be completed with words presented earlier in one of the lists you just read. However, other stems can be completed only with new words, that is, words that were not earlier presented. No plurals or proper names are allowed as completions.

What we are interested in is seeing whether people can avoid using the earlier-presented words when completing stems. Therefore, when a completion word comes to mind, you should not just say the word. Instead, you should check your memory to be sure that the completion word that has come to mind is not one of the words that was presented earlier. If it is, you are to complete the stem with a different word—one that was not earlier presented. It is important that you not give an earlier-presented word. If your completion word seems at all familiar, don't give it, but rather think of a different word. If you cannot think of a different completion, just leave the stem blank. It is better to leave the stem blank, than to use an earlier-presented word.

You will have 15 s to complete each stem. After 15 s, the computer will simply go on to the next trial. So remember, it is okay to let the time run out rather than complete a stem with a word you think may have been presented earlier. Use the time to try to think of a different completion—all the stems have multiple completions so thinking of a different word should be possible. But, if you can't think of a different word, don't be tempted to use the earlier-presented word—remember we are trying to see if people can avoid using the earlier-presented words. So stop and check each word before you give it as a response.

Inclusion Test Instructions

The following test is again a word-stem completion task in which some of the stems can be completed with earlier-presented study words and some only with new words. However, this time, we want to see how quickly you can complete the stems without worrying about whether your completion word was presented earlier or not.

Therefore, your task is to complete the stem as fast as you can with the first 5-letter word that comes to mind that fits the stem, without worrying about whether the word was presented earlier or not. Don't try to use memory because doing so will just generally slow you down. Rather, just complete the stem with the first word that comes to mind, doing so as rapidly as possible. You will have 15 s to complete each stem. After 15 s, the computer will simply go into the next trial. Again, no plurals or proper names are allowed as completions.

Experiment 2: Direct-Retrieval Instructions for Remember–Know

The next part of the experiment involves a word-stem completion task. You will be presented with the first three letters of a 5-letter word and your task is to complete the word. Your task is to use each stem as a cue to help you recall a word that was presented earlier in the experiment in either of the lists and to use that word as the completion for the stem. If you cannot recall an earlier-presented word, use the first word that comes to mind that completes the word stem. Plurals and proper names are not allowed.

After you have completed each word stem, your task is to decide whether each word is one you remember was presented earlier in the experiment, know was presented earlier in the experiment, or is new—that is, wasn't presented earlier in the experiment.

You are to respond that you "remember" a word as earlier presented if you consciously remember the circumstances under which you encountered the word earlier in the experiment. As an example, the word might bring back something you experienced when it was earlier presented, such as an image, or some personal significance of the word that you thought of when it was presented. You might remember the way it looked on the screen. In any of these cases, you should say "remember."

"Know" means that you know for a fact that the word was presented earlier in the experiment even though you don't remember any details about it. The word is just familiar in the experimental context. This is similar to seeing someone on the street, being aware that you know them, but being unable to establish any details about where you know them from. If you know the word was presented earlier in the experiment, but can't think of any details about it, say "know."

"New" means the word did not appear in either of the lists in the first part of the experiment. If the word is new, say "new."

So to summarize: You are to use the stem as a cue to help you recall an earlier-presented word. If you are unable to recall an earlier-presented word, complete the stem with the first word that comes to mind. You will have 15 s to complete each stem. After 15 s, the computer will simply go on to the next trial. After you have completed the stem, tell me whether the completion word is one you "remember," "know," or "new."

Appendix C

Multinomial Models Used to Fit Data From Experiments 1 and 3

Test item and parameter	Solution	Test item and parameter	Solution
Direct-retrieval model		Generate–recognize model	
Inclusion (inc)		Inclusion	
Long duration		Long duration	
$R_{long} * A_{long}$	target	$A_{long} * R_{long}$	target
$R_{long} * (1 - A_{long})$	target	$A_{long} * (1 - R_{long})$	target
$(1 - R_{long}) * A_{long}$	target	$(1 - A_{long}) * G_{inc} * R_{long}$	target
$(1 - R_{long}) * (1 - A_{long}) * G_{inc}$	target	$(1 - A_{long}) * G_{inc} * (1 - R_{long})$	target
$(1 - R_{long}) * (1 - A_{long}) * (1 - G_{inc})$	nontarget	$(1 - A_{long}) * (1 - G_{inc})$	nontarget
Short duration		Short duration	
$R_{short} * A_{short}$	target	$A_{short} * R_{short}$	target
$R_{short} * (1 - A_{short})$	target	$A_{short} * (1 - R_{short})$	target
$(1 - R_{short}) * A_{short}$	target	$(1 - A_{short}) * G_{inc} * R_{short}$	target
$(1 - R_{short}) * (1 - A_{short}) * G_{inc}$	target	$(1 - A_{short}) * G_{inc} * (1 - R_{short})$	target
$(1 - R_{short}) * (1 - A_{short}) * (1 - G_{inc})$	nontarget	$(1 - A_{short}) * (1 - G_{inc})$	nontarget
New		New	
G_{inc}	target	G_{inc}	target
$(1 - G_{inc})$	nontarget	$(1 - G_{inc})$	nontarget
Exclusion (exc)		Exclusion	
Long duration		Long duration	
$R_{long} * A_{long}$	nontarget	$A_{long} * R_{long}$	nontarget
$R_{long} * (1 - A_{long})$	nontarget	$A_{long} * (1 - R_{long})$	target
$(1 - R_{long}) * A_{long}$	target	$(1 - A_{long}) * G_{exc} * R_{long}$	nontarget
$(1 - R_{long}) * (1 - A_{long}) * G_{exc}$	target	$(1 - A_{long}) * G_{exc} * (1 - R_{long})$	target
$(1 - R_{long}) * (1 - A_{long}) * (1 - G_{exc})$	nontarget	$(1 - A_{long}) * (1 - G_{exc})$	nontarget
Short duration		Short duration	
$R_{short} * A_{short}$	nontarget	$A_{short} * R_{short}$	nontarget
$R_{short} * (1 - A_{short})$	nontarget	$A_{short} * (1 - R_{short})$	target
$(1 - R_{short}) * A_{short}$	target	$(1 - A_{short}) * G_{exc} * R_{short}$	nontarget
$(1 - R_{short}) * (1 - A_{short}) * G_{exc}$	target	$(1 - A_{short}) * G_{exc} * (1 - R_{short})$	target
$(1 - R_{short}) * (1 - A_{short}) * (1 - G_{exc})$	nontarget	$(1 - A_{short}) * (1 - G_{exc})$	nontarget
New		New	
G_{exc}	target	G_{exc}	target
$(1 - G_{exc})$	nontarget	$(1 - G_{exc})$	nontarget

Note. The models used for Experiment 1 were identical to these models, with the long-duration parameters becoming full-attention parameters and the short-duration parameters becoming divided-attention parameters. R = the parameter for consciously controlled process; A = the parameter for automatic process; G = the parameter for guessing process.

(*Appendixes continue*)

Appendix D

Parameter Estimates Across Instructional Conditions in Experiments 1 and 3 Calculated With the Direct-Retrieval and Generate–Recognize Models (Appendix C)

Condition and experiment	Parameter estimates					
	Direct-retrieval model					
Direct-retrieval condition						
Experiment 1	R_{full} .29	R_{div} .13	A_{full} .22	A_{div} .22	G_{inc} .45	G_{exc} .43
Experiment 3	R_{long} .42	R_{short} .21	A_{long} .30	A_{short} .30	G_{inc} .41	G_{exc} .39
	Generate–recognize model					
Generate–recognize condition						
Experiment 1	R_{full} .68	R_{div} .42	A_{full} .29	A_{div} .29	G_{inc} .45	G_{exc} .29
Experiment 3	R_{long} .70	R_{short} .47	A_{long} .42	A_{short} .42	G_{inc} .46	G_{exc} .37

Note. All analyses were conducted with the multinomial binary tree program (Hu, 1995; Hu & Batchelder, 1994). For all analyses reported here, the model was constrained so that $A_{full} = A_{div}$ (or $A_{long} = A_{short}$, in Experiment 3). Parameter estimates are shown only for conditions where the model fit the data from those conditions. R = the parameter estimate for consciously controlled process; full = full attention; div = divided attention; A = the parameter estimate for automatic process; G = the parameter estimate for guessing process; inc = inclusion; exc = exclusion.

Received October 25, 1996
Revision received June 16, 1997
Accepted June 22, 1997 ■