AMERICAN JOURNAL OF PSYCHOLOGY
March 1981, Vol. 94, No. 1, pp. 37-64

002-9556/80/0300-0037$02.80/0

Cognitive operations and decision bias in reality monitoring

Marcia K. Johnson
State University of New York
at Stony Brook

Carol L. Raye
Barnard College

Hugh J. Foley
State University of New York
at Stony Brook

Mary Ann Foley
State University of New York
at Stony Brook

In each of the three experiments, a reality monitoring task required subjects to discriminate between words they generated and words presented by an experimenter. Each of the experiments included a manipulation designed to affect the amount of external control over what the subject generated, with the expectation that the more a response is determined by external cues, the less the memory will include information about cognitive operations that took place when the memory was established. In general, increasing cognitive operations increased accuracy of reality monitoring. In addition, when subjects falsely recognized new items as old, they were much more likely to attribute the items to external sources than to internal sources. These findings were discussed primarily in terms of the role that cognitive operations preserved in memory may play in identifying the origin of information in memory. A comparison of memory for the occurrence of experimenter-presented and subject-generated items, regardless of correct identification of origin, extended the generation effect found by Slamecka and Graf in 1978 to information only covertly generated by the subject (Experiment 1), and to retention intervals as long as 10 days (Experiment 2). The results of Experiment 3 suggested that the generation effect may not necessarily appear in situations in which what is generated is essentially a meaningful response to what is perceived.

Reality monitoring refers to the processes involved in discriminating between perceptually derived and self-generated information in memory (Johnson, 1977). When people appear to confuse memories from these two sources, theoretical explanations have tended to attribute the errors to a degrading of information initially available, along with, perhaps, integrative processes, that serve to change the characteristics of the memory traces (e.g., Bransford & Franks, 1971; Loftus, 1975; Sachs, 1967). Johnson and Raye (in press, Note 1)

have instead proposed a model of reality monitoring that emphasizes the role of decision processes in reality monitoring, rather than changes in traces over time. The model postulates two basic types of decision processes: those based on proposed characteristic differences in the traces derived from perception and thought, and those based on reasoning about the specific content of memories, in light of past knowledge and metamemory assumptions.

With respect to the class characteristics, memories originating in perception should have more sensory information, time and place information, and more meaningful detail, while memories originating in thought should have more information about the cognitive operations (such as reasoning, search, decision, and imagery processes) that took place when the memory was established. These differences between externally and internally derived memories in average value along these dimensions or attributes could form the basis for deciding the origin of a memory. For example, a memory with a great deal of cognitive operations information and not very much sensory information could be judged to have been internally generated. More extended reasoning processes include considering whether the memory could have been perceived (or self-generated) given other specific memories or general knowledge. For example, a memory of a conversation with another person might correctly be attributed to a fantasy on the basis of the knowledge that you are not acquainted with the person. In addition, judgments might be expected to be affected by people's opinions or by "metamemory" assumptions about how memory works.

The present experiments provide evidence for two ideas incorporated in this model: (a) cognitive operations can serve as cues about the origin of memories, and (b) metamemory assumptions affect reality monitoring as revealed both directly through subjects' introspections and indirectly through decision biases operating in reality monitoring judgments.

EXPERIMENT 1

Slamecka and Graf (1978) reported the results of several experiments specifically directed at determining whether there is a memory advantage for self-generated information. Their general procedure was to present subjects with a stimulus and either to present a response (E-items) or to require the subject to generate a response (S-items) according to a rule and a specified first letter. The rule and first letter constrained the generated responses (e.g., *synonym-rapid-f* almost always produced the response *fast*) allowing Slamecka and Graf to counterbalance the assignment of items to conditions such that E and S items were the same across subjects. Although it was not the major focus of their research, they asked subjects to indicate the source of each item on an immediate recognition test in one of these experiments, and found, as we have (Johnson, Raye, & Durso, 1980; Raye & Johnson, 1980), that subjects could discriminate better than chance between experimenter-presented and subject-generated items. The present Experiment 1 yields additional evidence regarding variables that affect the amount of confusion between internally and externally generated memories, using procedures similar to Slamecka and Graf's.

We hypothesized that the procedure of cueing the subject with the first letter of the response might increase confusion between perceived and generated items by reducing the extent or salience of the cognitive operations involved in generation by specifying the response for the subject more exactly. In other words, the stronger or more specific the cue for a response, perhaps the more automatically the response is generated, reducing search and decision processes that later could be used as cues to the origin of the information. Hence, we compared subjects cued with the first letter of the response with subjects who were not cued with the first letter, expecting there would be better identification of origin without the first letter cues.

Primarily for generality, we used two types of items — one previously used by Slamecka and Graf, opposites (*hot-cold*), and a new class, category instances (*color-blue*), which we had previously used in another reality monitoring paradigm (e.g., Johnson, Taylor, & Raye, 1977; Raye, Johnson, & Taylor, 1980). We expected identification of origin in general to be lower for the opposites since generating an opposite for a word (e.g., *cold* in response to *hot*) should require very little in the way of search and decision procedures because there are so few candidates. Furthermore, because in the category case, the stimulus for each item (a category label) is a different class of item from the response (a category instance), specific confusion from stimuli (which are all externally presented) should have been minimized.

We also manipulated whether subjects responded overtly by writing down E and S words as they occurred or whether they were required to respond covertly by simply listening to E words and thinking about S words. With a different paradigm, we had previously found that overt responding increased confusion between perceptions and self-generated memories (Johnson, Taylor, & Raye, 1977), and we were interested in whether a similar effect would be found with the present paradigm. In addition, Slamecka and Graf's studies involved overt responding, and it was of interest to determine whether the advantage in memory of self-generated information that they and others had noted (Jacoby, 1978; Raye, Johnson, & Taylor, 1980) occurred when the self-generated information was only thought, but not expressed.

METHOD

There were three between-subject factors: Type of Material (category instances vs. opposites), Type of Expression (overt vs. covert), and Type of Cue (first letter vs. no first letter). Origin of Response (experimenter or external vs. self or subject-generated) was a within-subject factor.

Category lists were made up of 36 category names and common instances (e.g., *color-blue*). Opposites were 36 adjectives and an appropriate opposite (e.g., *hot-cold*). For the category list, the 36 items were randomly assigned to E items and S items, 18 of each. The presentation sequence was determined by randomly assigning E items to three groups of 6 items each and randomly assigning S items to three groups of 6 items each. E and S blocks of 6 items were then alternated. A second set of materials was created by changing each E item to an S item and vice versa. Both sets were then recorded on tape at a 6-sec rate. Each block was preceded by a signal that indicated whether the subject was to "listen" or "generate" for the next block of items. Half of the subjects were randomly assigned to one set and half to the other. An analogous procedure was followed for the opposite conditions.

Subjects participated in small groups of up to four people and signed up for sessions. Sessions were randomly assigned to conditions in randomized blocks of the eight between-subject conditions. Each condition was run at least four times. The subjects were Stony Brook undergraduate volunteers. Conditions were run until there were at least 12 subjects per condition.

Material sets were randomly assigned to sessions. There were two experimenters, each assigned an equal number of the randomized blocks of the experimental conditions. For all conditions, the procedure was first illustrated with a practice tape consisting of two E item trials and two S item trials with practice items appropriate for the condition. Subjects covered previous responses with a mask. The response sheet included the first letter of both experimenter and self-generated items for the first-letter conditions. Covert subjects indicated with a check mark or a zero whether or not they had heard the response in the case of E items and whether or not they had generated a response in the case of S items. Overt subjects wrote down both E and S responses. All subjects were instructed to try to generate very familiar and common responses, ones they thought most people would give in this situation. They were not warned about a memory test in advance. After the response sheets had been collected, subjects were asked to recall all of the responses that they could in any order. The instructions emphasized that both the responses given on the tape and the ones they generated were equally important. Four min were allowed for recall. Next, a new test sheet was distributed for the cued recall and origin test. No first-letter cues were given for this test. Each stimulus was read to the subjects in a random order and the subjects were asked to write down the response that went with it — either the one provided or the one they generated. In addition, they were also asked to indicate whether each response had been given on the tape or whether they had generated it themselves. The origin test proceeded at a 7 sec rate.

RESULTS[1]

Acquisition

While failures to respond were quite rare, as would be expected, there were more omissions for the subject-generated items (mean = .51) than for the experimenter-generated items (mean = .09), $F(1, 88) = 23.16$, $MS_e = .36$. No other variables affected omissions.

For E items, nontarget responses written by overt subjects were extremely rare (mean = .10) and were not affected by the experimental variables.

For overt subjects, the mean number of generated responses that were not targets — that is, that were not the response selected to be the E responses when that item was assigned to the E, rather than the S, condition — were: opposites-no first letter, 3.00; opposites-first letter, 1.00; category instance-no first letter, 10.58; category instance-first letter, 3.33. Examples of these types of responses were *wrong* (rather than *left*) in response to *right*, or *cat* (rather than *dog*) in response to *animal*. These alternative target responses happened much less often in the case of opposites than in category items, $F(1, 44) = 116.33$, $MS_e = 2.54$, and less when the response was cued with the first letter, $F(1, 44) = 101.22$. As might be expected, because the number of appropriate alternative responses is so much greater in the case of category instances compared with opposites, the reduction in alternative-target responses produced by the first letter cue was much greater in the case of category items, $F(1, 44) = 32.60$.

For the following analyses of free and cued recall, both targets and alternatives were considered correct. (Corresponding analyses of only target responses will be discussed briefly later but yielded comparable statistical outcomes in all important particulars.) For the covert conditions, it was assumed that targets or target alternatives produced on free and cued recall had also occurred during acquisition, although, of course, there was no direct way of verifying this. Thus, if any subject recalled an item on the free recall test, for

example, that was an appropriate response to an S item, it was considered an alternative target. A comparison of items designated alternative targets by this criterion with actual responses at acquisition for overt subjects indicated that it was very unlikely for a subject to suddenly produce *dog* or *left* on the free recall test if these had not been produced at acquisition. Thus it seemed reasonable to assume that items recalled by covert subjects also had been thought earlier at acquisition. In the analyses reported below, both absolute number and proportions were analyzed and generally gave the same results. Therefore, only the proportions are reported and discrepancies in analyses of these two dependent variables are noted.

Free recall

The measure here is free recall of targets and target alternatives expressed as a proportion of the total number of responses produced (or, for covert subjects, checked) during acquisition. There were three major effects. Category items (.38) were recalled better than opposites (.24) were recalled, $F(1, 88) = 54.63$, $MS_e = .02$. Overt responding (.36) produced higher recall than covert responding did (.26), $F(1, 88) = 22.22$. And the items the subjects generated themselves (.38) were recalled better than the experimenter items (.26) were recalled, $F(1, 88) = 30.17$, $MS_e = .02$. There were no other significant effects. It is especially interesting that the advantage of self-generated over experimenter-presented items was not influenced by whether or not the subjects responded overtly, and held separately for both overt, $F(1, 44) = 12.12$, and covert, $F(1, 44) = 18.30$, conditions alone: For overt, S = .41, E = .31 and, for covert, S = .33 and E = .20.

Cued recall

On the cued recall test, subjects were given the stimulus and asked to produce the response that had gone with it at acquisition. In general, cued recall was high in all conditions; however, there were some significant differences as a consequence of the experimental manipulations. The mean proportion cued recall was .96 and .99 for opposite E and S items, respectively, and .83 and .97 for category E and S items, respectively. Cued recall was higher for opposites compared to category instances, $F(1, 88) = 23.01$, $MS_e = .01$. This could very likely reflect guessing rather than remembering, since the probability of being correct with a good guess would be so much higher for opposites than for category instances. In addition, there was an interaction of type of item with origin, $F(1, 88) = 23.01$, $MS_e = .01$. Both opposites and categories showed an advantage for self-generated over experimenter-presented information. However, the advantage was greater in the case of the category items; this very likely reflects a ceiling effect for opposites. For category items alone, the E vs. S comparison was significant, $F(1, 44) = 50.29$, $MS_e = .01$. For the opposites alone, the E vs. S comparison was significant in the analysis of proportions, $F(1, 44) = 9.05$, $MS_e = .003$, but not in the analysis of number correct.

It should also be emphasized that the advantage of self-generated over experimenter-presented information was similar for overt and covert responding and was significant for both cases separately: for overt responding, S(.98) vs. E(.90), $F(1, 44) = 37.68$, $MS_e = .005$ and for covert responding, S(.98) vs. E(.89), $F(1, 44) = 23.89$, $MS_e = .01$.

Identification of origin

The proportion of items recalled on the cued recall test that were also attributed to the correct source are shown in Table 1. Since several variables interacted with type of materials, we will examine the results for opposites and category instances separately. For category items, identification was superior, as predicted, for the no-first letter conditions (.94) compared with the first-letter conditions (.88), $F(1, 44) = 7.30$, $MS_e = .01$. Second, identification was better for subject than it was for experimenter (E) items, $F(1, 44) = 14.47$, $MS_e = .01$. In addition, there was an advantage for subjects in the overt responding condition compared with subjects in the covert responding condition $F(1, 44) = 7.33$, $MS_e = .01$; and there was an origin by expression interaction $F(1, 44) = 4.90$. Overt expression appeared to help experimenter-presented items more than it helped self-generated items; however, from Table 1, it seems that this interaction might have been produced by a ceiling effect in the case of subject-generated items.

Within the opposites, there was no significant effect of first-letter cueing. Identification was better for experimenter-presented items than it was for subject-generated items, $F(1, 44) = 6.61$, $MS_e = .01$,

Table 1. Proportion-coreect identification of origin, Experiment 1

	Opposites		Category instances	
	Overt	Covert	Overt	Covert
E items	.80	.71	.92	.81
S items	.76	.64	.96	.95

and overt responding was superior to covert responding, $F(1, 44) = 10.70$, $MS_e = .03$. In general, performance was lower for opposites than it was for category instances $F(1, 88) = 81.98$, $MS_e = .02$.

Methodological considerations

The probability that the subject initially generated the desired targets varied with condition, as noted previously. In the above analyses, alternative responses were treated as targets. However, it is possible that the alternatives differed from the targets in some way that influenced their recallability or their discriminability from external items. Consequently, additional analyses were conducted using only the actual targets generated by each subject. These analyses produced results that were entirely comparable to those already reported. For example, in the overt conditions, the proportion free recall of S and E targets was .42 and .30, respectively, $F(1, 44) = 12.88$, $MS_e = .02$. The mean proportion cued recall was .95 and 1.00 for opposite E and S items, and .86 and .99 for category E and S items, respectively. The $F(1, 44)$ values for the main effect of source and the main effect of materials were 42.23, $MS_e = .01$, and 11.42, $MS_e = .01$, respectively. Finally, for the category items, identification of origin was superior for the no-first letter conditions (.94) compared with the first-letter conditions (.87), $F(1, 44) = 9.57$, $MS_e = .01$.

Another issue concerns the procedure of administering more than one test to a subject. Ideally, separate groups of subjects should have received the free recall, cued recall, and identification-of-origin tests. However, the cued recall and identification-of-origin test followed the free recall test, and thus the potential effects of the free recall test on subsequent measures should be considered. A generation effect was found on the cued recall test; cued recall of S items was greater than cued recall of E items. Prior free recall quite likely weakened the magnitude of this effect because some E items were generated during recall and thus derived whatever benefit obtains from being generated (assuming the effect of generating an E-item once is greater than the effect of a second generation of an S-item). On the identification-of-origin test, prior free recall quite likely would create confusion for some items because E-items that were recalled received a generation that could be confused with the earlier presentation. Thus, the overall level of reality monitoring might have been depressed. In addition, this potential confusion might influence more category items than opposites because more category items were produced on the free recall test. As a consequence, the advantage of category items over opposites on identification of origin, while clearly significant, was perhaps smaller in magnitude than it might have been without the recall test. Finally, neither free nor cued recall was influenced by whether or not information had been generated with first-letter cues; therefore, identification-of-origin comparisons involving this variable can probably be interpreted in a straightforward fashion.

DISCUSSION

According to the reality monitoring model, increasing automaticity in responding ought to reduce information regarding cognitive operations. Conversely, if a response requires somewhat more effort or search, stored information about these operations should become a potential cue for identifying the source of a memory. Two aspects of the present data are consistent with this idea. First, in the category condition, generating instances of categories to fit first letter cues, compared with generating instances of categories without first-letter cues, resulted later in lower correct identification-of-origin. This is consistent with the proposition that as responses are more related to ongoing external stimuli, and as they are more controlled by external cues, these responses will be more difficult to discriminate from external events. Upon reflection, it is reasonable that the first letter manipulation would have had less effect in the case of opposites; even without the first letter cue, the response is almost completely determined by the combination of the orienting instructions to give opposites and the stimulus, and thus requires little in the way of search and/or decision processes. Furthermore, we know that providing first-letter cues did not simply depress performance, because this manipulation did not affect free or cued recall. (The only other significant effect of first letter cues was beneficial; they reduced stimulus intrusions in free recall in the case of opposites.)

Second, the materials by origin interaction can be understood by considering the role of cognitive operators as cues to the origin of information. The identification of origin of E items as compared with S items was greater in the case of opposites whereas S rather than E items were more easily identified in the case of categories. As will be shown in Experiment 2, there is an overall bias to identify an item as E rather than S generated, for both opposites and category instances. This bias should be most likely to operate on items for which there is minimal cognitive operation information. Thus, when unsure, subjects should have been more likely to say E than S and this tendency would raise the identification-of-origin score on E

items. Assuming then that generating opposites produced less operation information than did generating category instances, this bias could have produced the higher identification scores on E items in the case of opposites.

An aspect of the present data that presents something of a puzzle is the superior identification of origin in the overt as compared with the covert conditions. It would seem that writing down both E and S items would *reduce* the discriminability between the two by increasing the overlap in similar information stored with each event, compared with simply listening to E items and silently generating S items (see Johnson, Taylor, & Raye, 1977; Experiment 2). One explanation of this finding takes into account the possibility that writing the item down is based on mental repetitions or re-generations of the initial events which are, of course, different in the case of E and S items. That is, as the subject writes down an E item, it is done with reference to the original external experience, and thus the memory of the external experience is being rehearsed (or the initial experience is extended). In contrast, as the subject writes down an S item, it is done with reference to the original internal event, and thus the internal event is being rehearsed. This reasoning is consistent with the finding that repetitions of different external and internal events increased their discriminability (Johnson, Raye, & Durso, 1980).

Another important finding of the present study was that the generation effect — that is, the superior memory for self-generated as compared with experimenter-presented information — was as great in both free and cued recall when the subjects only thought about, but did not overtly express, their responses. This outcome supports the idea that the locus of the advantage is in the generation process itself and does not, for example, depend upon overt expression of the generated item and consequent sensory components (e.g., hearing one's own voice or seeing one's own handwriting), or on some combination of generation and sensory components.

EXPERIMENT 2

Experiment 2 involved reality monitoring decisions about E and S items of three types: (a) opposites, (b) high frequency or typical category instances, and (c) low frequency or less typical category instances. Both E and S items were cued with the first letter and were written by the subject at acquisition. First letter cues were used to control the amount of search that would be required to generate category instances. The opposites and typical category instances were equivalent to the materials used in Experiment 1. Typical category instances consisted of the primary response in the Battig and Montague (1969) norms. Less typical items were selected such that they were also quite common responses but so that they were somewhat lower in the norms. The expectation was that the greater effort or search (or extended retrieval) for the less typical items would increase the chances that information about cognitive operations would be available and hence increase correct identification of origin compared with the more typical instances.

Experiment 2 included a number of other changes in design that we thought would be informative. As each E or S item was written down at acquisition, subjects were asked to indicate the probability that they would be able to remember it. Subjects were also asked to make an overall prediction about future memory performance after the acquisition procedure ended. These predictions provided information about potential decision rules that might be relevant in reality monitoring. Also, new items, not previously presented or generated, were included on the identification-of-origin test. This permitted an assessment of whether subjects make "false positives" in an unbiased manner, or whether the decision rules used were more likely to produce an "E" or an "S" response. In addition, at the time subjects were tested on identification of origin, confidence ratings were taken as a potentially more sensitive measure of origin identification than percentage correct.

Finally, the identification-of-origin test was given after a 10-day retention interval. According to our working model, decision processes in reality monitoring are applied to information that is preserved in memory over substantial retention intervals, and does not simply reflect a very transistory availability of "surface" information. We were able to determine, as well, whether the superior recognition of self-generated compared with experimenter-presented information (regardless of correct identification of origin) is a phenomenon that will be observed at long intervals.

METHOD

In phase 1, subjects saw stimuli and first letter cues for 48 items. In addition, the relationship between each stimulus and response was indicated (category instance or opposite). Half the stimuli were category names and half were opposites. For half the items of each type, the responses were given; for the remaining items, the subjects were asked to fill in an appropriate response. Within the category items, half of the instances were the most common instance (Battig & Montague, 1969) of the category /(category-high) and half the instances were less common instances (category-low). Some examples of category-high items are *fruit-apple, sport-football,* and of category-low items are *animal-pig, flower-lily.* As you can see, the category-low targets were still relatively common instances of the category. Two lists were constructed which were the same except that the assignment of items to externally presented (experimenter-presented) and subject-generated categories was reversed. The stimuli and the responses, where given, were typed on a study sheet. The list began with several filler items. Critical items were randomly distributed with the restriction that items from the various conditions be equally distributed throughout the sequence. On a separate response sheet, the stimuli and first letter cues were typed, and it was the subject's task to write down the response that had been given or to generate one which started with the letter indicated. Subjects were encouraged to give common responses. After subjects wrote down each response, they indicated how confident they were that they would later be able to remember that word by circling a number on a percentage scale that ranged from 0 to 100 in increments of 10. As well as providing information about whether subjects' estimates of the memorability of items was influenced by origin (externally or internally generated), this cover task also served the purpose of insuring attention to each item. Subjects were paced through this first phase of the experiment at a 7-sec rate by having them move masks down the study page. At the end of this phase, subjects were asked to make an overall judgment about whether E or S items would be easier to remember.

During a 10-day retention interval, a separate recognition list was constructed for each subject. The recognition list consisted of all E items and S items from phase 1 plus 12 new instances of categories (6 selected according to the category-high criteria and 6 selected according to the category-low criteria) and 12 new instances of opposites. The old and new items were randomly intermixed according to a pattern which guaranteed that items from each condition were distributed equally across the test order. In the few instances where the subject had not been able to generate an appropriate response during phase 1, the abbreviation NR appeared in place of the item. Each item was printed on a computer card, along with the possible responses E, S, or N (for external, self-generated or new) and the numbers 1 to 5 to indicate a confidence rating for each response. On this scale, 1 indicated a guess and 5 indicated very high confidence in the accuracy of the response. The recognition lists were randomly ordered and printed by computer. During the recognition test, each subject received the appropriate deck of cards. The recognition instructions included an explanation of the NR cards. The subjects were paced through the recognition test by having them turn cards at a 5 second rate in response to an auditory signal.

The subjects were 34 Barnard College students all of whom participated simultaneously.

RESULTS

In order that both opposite and category items would be treated as equally important by the subjects, there were twice as many opposite items as there were category-high or category-low items. Therefore, where appropriate, measures of performance on opposite items were divided by two in order to be able to compare performance across the three item conditions.

Acquisition

With rare exceptions (mean = .97), subjects produced the expected targets for opposites. As in Experiment 1, there were more target alternatives produced for category instances, with, as might be expected, a higher number of alternatives produced in the category-low case (mean = 2.53) than in the category-high case (mean = 1.23). Examples of alternatives commonly given are *lilac* instead of *lily, Ohio* instead of *Oregon*, and *parrot* instead of *pig*. Since individual test lists were generated for each subject, target-alternatives were, as in Experiment 1, treated as targets in the following analyses.

Recognition

Two types of analyses were performed on the recognition scores: in one set, the raw data were used and in the other set, proportions were used to correct for small differences in no-response rates. The results of these two sets of analyses were entirely comparable, and only the proportions will be reported.

The mean proportions of misses (calling an old item new) are shown in the top half of Table 2. In general, there were more misses on external than on subject-generated items, $F(1, 33) = 14.12$, $MS_e = .03$. Thus a generation effect was apparent even after 10 days. Misses were not affected by item type, nor did item type interact with origin. The false positives are also shown in Table 2. The total false positives differed as a function of type of item, $F(2, 66) = 55.36$, $MS_e = .01$.

Table 2. Mean proportions of misses, false positives, and correct identification of origin, Experiment 2

Measure	Opposites	Category-high	Category-low	Mean
Misses				
E words	.36	.37	.42	.38
S words	.26	.29	.33	.29
Total	.31	.33	.38	
False positives				
E responses	.30	.18	.11	.20
S responses	.24	.08	.07	.13
Total	.27	.13	.09	
Correct identification of origin	.57	.55	.67	.58

Subsequent analyses indicated that there were more false positives on new opposites than there were on either new category-high items, $F(1, 33) = 52.43$, $MS_e = .01$, or new category-low items, $F(1, 33) = 91.24$, $MS_e = .01$, and more false positives on new category-high than on new category-low items, $F(1, 33) = 6.47$, $MS_e = .01$.

Identification of origin

False positives broken down by the type of false positive response, E or S, are shown in Table 2. Subjects were less likely to incorrectly call a new item a self-generated item than they were to call a new item an externally-presented item, $F(1, 33) = 7.72$, $MS_e = .03$.

The mean proportions of correct identification of origin for each type of item are also shown in Table 2. Given that subjects recognized an item as old, their ability to discriminate the source of the item varied as a function of the type of item, $F(2, 66) = 4.47$, $MS_e = .03$. As is obvious from Table 2, the main effect was produced by the predicted greater ease of identifying the origin of category-low items vs. opposite, $F(1, 33) = 6.92$, $MS_e = .02$, and category-low vs. category-high, $F(1, 33) = 8.03$, $MS_e = .03$. With this measure, the greater accuracy for category-high as compared with opposites found in Experiment 1 was not evident, perhaps as a consequence of the much longer retention interval (10 days as compared to a few minutes). The confidence ratings reported in the next section were more sensitive to differences in materials.

While the category-low items had the highest identification-of-origin scores, they were also the items most likely to have been something other than the expected targets. It is possible that subjects generated a set of idiosyncratic items that were more discriminable from the E items in, say, the category-low compared with the category-high condition. Table 3 shows a summary of performance on S items separately for targets and alternatives. As can be readily seen, the experimental manipulation of category frequency was not effective for items that were not the targets (.43 vs. .37). On the other hand, for targets, the manipulation worked as expected, with the origin of lower frequency items more readily identifiable (.38 vs. .54). Assuming the alternatives are likely to be highly available responses for an individual (and thus involve few cognitive operations to produce) idiosyncratic responses should look like high frequency targets, and they do. Thus, rather than exaggerating any differences, the higher rate of idiosyncratic responses in the low frequency case probably weakens the differences reported here.

Table 3. Proportion of hits and misses for targets and alternatives generated at acquisition, Experiment 2

	Hits		
	Source correctly identified	Source incorrectly identified	Misses
Idiosyncratic target alternatives			
Category-high	.43	.21	.36
Category-low	.37	.23	.37
Expected targets			
Category-high	.38	.34	.28
Category-low	.54	.18	.28

Confidence ratings of origin identification

The mean confidence ratings given to items correctly identified as old are shown in Table 4. The first three columns show the ratings for items whose origin was correctly identified, and the other three columns give the confidence ratings for items correctly identified as old but misidentified with respect to origin. In general, the items where the origin was correct received higher confidence ratings than the items where the origin was incorrect, $F(1, 33) = 12.01$, $MS_e = 2.53$. (Compare the third and sixth columns.) These data, like the proportion correct identification of origin reported in the last section, again indicate that subjects were discriminating subject from experimenter items even after 10 days. There was also a significant interaction between materials and correct/incorrect identification, $F(2, 66) = 4.43$, $MS_e = 1.37$. For all three types of materials, there was a higher confidence associated with correct identification of origin than with incorrect identification of origin, but this advantage was greatest in category-low items, next greatest in category-high items and least for opposites, as would be expected on the basis of the relative amounts of cognitive operation information they would produce during generation. The F values for correct/incorrect comparisons for each of these groups separately were, $F(1, 33) = 11.92$, 3.78, and 3.23 for category-low, category-high, and opposite, respectively. As mentioned above, compared with the proportion-correct identification of origin measure, these confidence ratings appear more sensitive and should prove useful in future studies, especially those involving long retention intervals. Subjects expressed higher

Table 4. Mean confidence ratings for hits, Experiment 2

	Correct-origin identification			Incorrect-origin identification		
	E responses	S responses	E & S combined	E responses	S responses	E & S combined
Opposites	2.73	3.26	3.00	2.75	2.95	2.85
Category-high	3.05	2.98	3.02	2.28	2.74	2.52
Category-low	2.98	3.94	3.46	2.09	2.84	2.47

confidence when identifying an item as self-generated than when identifying an item as experimenter-presented, whether or not they were correct, $F(1, 33) = 11.47$, $MS_e = 2.00$.

Interestingly, higher confidence ratings for S responses than for E responses appeared only for actual old items, and not for new items incorrectly identified as old, as can be seen in Table 5, $F(1, 33) = .27$, $MS_e = 2.21$. The confidence ratings associated with E and S items that the subject incorrectly identified as new also did not differ, as shown in the lower portion of Table 5, $F(1, 33) = 1.87$, $MS_e = 1.48$. Finally, confidence ratings on false positives were influenced by the materials, $F(2, 66) = 33.96$, $MS_e = 1.43$. From Table 5, this primarily reflects the higher confidence ratings given to opposites. These data support the notion that one of the cues for recognition is the readiness with which some specific response occurs to an item (Raye, 1976).

Subject predictions

At acquisition, subjects indicated on a percentage scale the likelihood they would later remember each item. The overall average judgment was 62.35, and the judgments did not vary as a function of either type of material or of origin of the item. However, at the

Table 5. Mean confidence ratings for errors, Experiment 2

Measure	Opposites	Category-high	Category-low
False positives			
E responses	2.54	1.53	1.19
S responses	2.79	1.20	.96
Misses			
E words	2.59	2.73	2.72
S words	2.30	2.30	2.74

end of the first phase of the experiment, the subjects were asked whether they thought the E words or the S words would be easier to remember or whether they would be of equal difficulty. The number of subjects giving each response was: S easier than E (16); E easier than S (1); no difference (14); no response (3). Thus, with a question directly asking for a comparison of externally presented items with self-generated items, those subjects who believed there would be a difference clearly expected the self-generated information to be easier to remember ($n = 17$, $z = 3.39$). In contrast, when each item was rated individually as it was experienced, the ratings were probably more controlled by semantic properties of the items themselves and not by their source. The difference in performance on these two indices points out that metamemory judgments are probably affected by a number of factors, and which of these will predominate (and thus perhaps control processing time or effort) depends on the situation.

To summarize the results of Experiment 2: in simple recognition, S items were at an advantage over E items even after 10 days. Subjects were able to identify the origin of old information after 10 days, as indicated by higher confidence ratings given to items recognized and correctly identified with respect to origin, compared with those recognized but not correctly identified. Furthermore, the difference between the confidence ratings given to correctly attributed and to incorrectly attributed items was greatest in the category-low condition, next for the category-high, and least for opposites, consistent with the relative amounts of cognitive operations the S items should have produced at acquisition. Finally, a bias in attribution was found: if subjects incorrectly identified a new item as old, they were more likely to say it was presented than that they generated it. An interpretation of this bias that is consistent with the idea that cognitive operations serve as cues to identification of origin will be presented following Experiment 3.

EXPERIMENT 3

In Experiments 1 and 2, S and E items were cued by different other items (e.g., category labels), and were not cued by each other. A common situation, for example, in conversations, is one in which self-generated events occur cotemporally (e.g., as a response) with external events from which one might later want to distinguish them. Experiment 3 approximated this situation by having subjects generate responses to E items. We again attempted to vary the amount

of cognitive operations involved in producing responses, this time not by varying the cue, but by varying the instructional set given to the subject. Some subjects were instructed to generate related responses to the E items and some subjects were instructed to generate unrelated responses. We expected that generating something not related to ongoing events should require more cognitive processing than generating something related would require. Furthermore, there were two types of related conditions: Some subjects were told to generate a word that had some meaningful relationship to each externally presented word; other subjects were told to give a word that started with the same first letter as the word they heard. Responses based on meaning should be more available; this is a more natural way of responding to the verbal environment that capitalizes on previously established associative relationships, and responses should be produced more automatically. Coming up with a word with the same first letter should be a more novel response to the word and hence may require somewhat more search and/or decision processes.

METHOD

Design

All subjects were read a list of 30 common, highly concrete English nouns, and were asked to generate a word in response to each word they heard. Three independent groups differed in the instructions they received about generating these words. Subjects in the first-letter condition were told to give a word which started with the same first letter as the word they heard (e.g., *apple-airplane*). Subjects in the related condition were told to give a word that has some meaningful relationship to the stimulus word (e.g., *apple-earth*). The example was purposefully chosen so that it did not illustrate a particular sort of relationship such as superordinate, part-whole, etc. in order to avoid a set effect. Subjects in the unrelated condition were instructed to give a word which was not related to the stimulus item (e.g., *apple-clown*). Immediately following this phase of the experiment, the subjects were asked to recall as many of both types of words (experimenter-presented or E words and subject-generated or S words) as they could. They were not asked to identify the origin of the words they recalled. After a week's delay, the same subjects were given a surprise recognition test consisting of experimenter-presented and subject-generated words and new words. They were at this time asked to indicate the origin (E, S, or N) of each item.

Materials and procedures

All subjects were tested together. Instruction booklets were passed out in randomized blocks of the experimental treatments. The instructions did not include any warning about possible memory tests. The 30 words were then read in random order at approximately a 10 sec rate. Subjects in all conditions first wrote the experimenter's word down on a sheet of paper and then, next to it, wrote the word they generated. Following this phase, subjects were instructed to turn over the sheet of paper on which they had been writing and to put their name on a new sheet of paper. Including irrelevant conversation, there was an interval of approximately 3 min between the end of acquisition and the beginning of recall. For the recall task, subjects were instructed to write down as many of the words — both the ones given by the experimenter and the ones they had generated — as they could. They were also told that they could recall words in any order, both types of words "counted equally," and that it was important that they try to remember as many of both types as possible. Approximately 5 min were allowed for recall.

Each subject's recognition list consisted of 45 words, a random selection of 15 of the E words, the 15 S words which did not correspond to the selected E words, and 15 new words selected from the same class of materials as were the original E words. There were two random patterns used for selecting among the various types of words, each used for half of the subjects. In the rare instances when a selected S word also was an E word or a new word, replacement items were chosen from the subject's protocol so that the final recognition list was comprised of items of unambiguous origin. The 45 words were then typed in random order with the letters *E*, *S*, and *N* typed next to each word. The final recognition lists were randomized and typed by computer.

The subjects were paced through the recognition test at a 5-sec rate per item using a mask to prevent them from glancing ahead at the list. They were instructed to circle the correct letter indicating whether they thought the item was one that the experimenter had given, one they had given, or a new item.

Subjects

The subjects were Stony Brook college students enrolled in a course on experimental methods. The data for those students who did not attend class for the second session were discarded. In addition, several other complete protocols were randomly discarded in order to equalize the *n*s at 16 subjects per condition.

RESULTS

Recall

The mean recall for each type of word in each of the three conditions is shown in Table 6. There was a main effect of instructional condition $F(2, 45) = 9.96$, $MS_e = 12.44$. Subsequent analyses indicated that the first-letter condition recalled fewer words than either the related condition or the unrelated condition, $F(1, 30) = 22.36$, $MS_e = 10.92$ and $F(1, 30) = 7.39$, $MS_e = 12.21$, respectively. The related and unrelated conditions did not differ significantly in overall recall, $F(1, 30) = 2.64$.

Table 6. Mean recall, Experiment 3

	Related	First-letter	Unrelated
E words	14.94	8.94	8.62
S words	13.19	11.38	16.44
Total	28.13	20.32	25.06

In the overall analysis, there was also an interaction of instructional condition and origin, $F(2, 45) = 35.40$; subjects in the related condition recalled more experimenter words than subject words, $F(1, 15) = 15.64$, $MS_e = 1.57$, while subjects in the unrelated condition recalled more subject words than experimenter words, $F(1, 15) = 60.42$, $MS_e = 8.08$. Like the unrelated subjects, the first-letter subjects also recalled more subject words than experimenter words, $F(1, 15) = 8.01$, $MS_e = 5.93$, but the difference in recall of the two types of words was greater in the unrelated condition, $F(1, 30) = 16.49$, $MS_e = 7.00$.

Recognition

A miss is calling either an E item or an S item "new." The first two rows of Table 7 show the mean misses as a function of each instructional condition and source of words. In the overall analysis of misses, there was an Instruction × Origin interaction, $F(2, 45) = 7.43$. Subsequent analyses indicated that for first-letter and unrelated conditions, there were more misses on E items than there were on S items, $F(1, 15) = 5.00$, $MS_e = 2.50$, and $F(1, 15) = 13.39$, $MS_e = 4.31$. The difference between E and S items in the case of related subjects was not significant, $F(1, 15) = 1.54$.

Table 7. Mean number of recognition errors and mean proportion of hits for which the origin was correctly identified, Experiment 3

Measure	Related	First-letter	Unrelated
Misses			
E words	3.50	3.56	4.62
S words	4.19	2.31	1.94
False positives			
E responses	2.62	4.81	4.38
S responses	1.00	1.50	1.25
Proportion correct identification of origin	.62	.68	.74

The mean total false positives (saying E or S to a new item) were 3.62, 6.31, and 5.63 for related, first-letter and unrelated conditions respectively, $F(2, 45) = 3.54$, $MS_e = 4.41$. The first-letter and unrelated conditions did not differ significantly, and the related condition had fewer false positives than either the first-letter, $F(1, 30) = 6.04$, $MS_e = 4.79$, or the unrelated, $F(1, 30) = 5.47$, $MS_e = 2.93$, condition.

Identification of origin

False positives are shown separately for E and S responses for each of the three conditions in Table 7. Overall, subjects were more likely to say E than S to a new item, $F(1, 45) = 49.85$, $MS_e = 3.48$, and this bias did not interact with instructional condition.

For those words correctly recognized as old, the proportion correct identification of origin was computed for each subject. The mean proportion correct identification of the source of an item was .62, .68, and .74, for related, first-letter and unrelated, respectively, $F(2, 45) = 4.39$, $MS_e = .01$. Thus the pattern of results suggests that as cotemporal thoughts become more related to perceptual events, the likelihood of later confusing their memory representations increases. However, while the related subjects showed the least ability to identify the origin of an item, they were still significantly above a chance score of 50%, $t(15) = 2.17$. It should be emphasized that the recall and recognition results argue against the idea that the related subjects simply remembered less in general. In fact, on recognition, they made significantly fewer false positive errors.

While our preferred interpretation attributes the relatively low reality monitoring performance in the related condition to the fact that the generated responses were closely related to external items, it should be noted that the related condition produced the highest recall of E items on the prior free recall test. Cues in memory from these test-context generations might then have made E items seem more like S items. If this were the major source of the low discriminability between E and S items, we would expect to see a particularly pronounced tendency to call E items S items in the related condition. Table 8 shows the proportion of E items that were recognized as old but attributed to the wrong source, along with the proportion of S items that were recognized as old but called E items. There was a significant interaction between instructional condition and type of misidentification, $F(2, 45) = 7.92$, $MS_e = .04$. The manipulation of instructions had relatively little effect on subjects' ability to identify the source of E items. Contrary to the idea that subjects in the related condition were especially likely to confuse test generations of E items with the target memories, these subjects were no more likely than other groups to call E items S. The instructional variable had its major impact on S items. As predicted by the reality monitoring model, the more cognitive operations it took to establish a memory, the less likely it was to be attributed to external sources. Thus the related subjects were not more likely to call E items S but were more likely to call S items E.

Table 8. Mean proportion of hits for which the origin was misidentified, Experiment 3

	Related	First-letter	Unrelated
E words called S	.26	.36	.32
S words called E	.52	.31	.20

Reported use of cues of various types

Following the identification of origin test in Experiment 3, subjects were asked to describe how they were able to identify the source of various items. Interestingly, subjects tended to mention sensory cues more often with respect to E items ("I differentiated words which you said by remembering your pronunciation," "I could visualize your saying it," or "The words which the E stated were remembered in her voice"). Cognitive processing, additional information, and semantic content were mentioned more in conjunction with S items ("When I was very sure [about my words] I could remember I had a very specific reason for making the association. If the word [only] seemed familiar, I would say that it was the experimenter's word," "I made the decision by knowing what my train of thought was during the exercise," "Sometimes the words I chose went together with a certain scene, i.e., pond, cloud, tree, etc. And, when I saw the words again I tried to remember if they fit in any of the images I had").

In summary, the recall and recognition data were for the most part consistent with each other and relatively straightforward. For both immediate recall and recognition memory after a week, memory for subject-generated information was superior to memory for external information in both the first-letter and unrelated conditions. In contrast, subjects who generated semantically related responses remembered slightly but significantly more external items on the immediate recall test and slightly but not significantly more external items when given the recognition test a week later.

As expected, correct identification of the origin of items was influenced by the degree to which the internal and external events were related. The poorest reality monitoring occurred when the external and internal events were highly related semantically. Consistent with Experiment 2, Experiment 3 also yielded evidence about a bias that undoubtedly influences decision processes in reality monitoring. When a completely new item was thought to be old, the subjects in all three conditions were about three times more likely to attribute it to an external source than to say they generated it themselves.

GENERAL DISCUSSION

The generation effect

As mentioned previously, Slamecka and Graf's major interest was comparing memory for E and S items, and they consistently found superior recall and recognition for subject-generated items across a number of different types of items (opposites, same category membership, synonyms, rhymes, associates) and variations in procedure (e.g., incidental vs. intentional instructions). The present experiments extend the phenomenon to category instances, covert generations, and considerable retention intervals (see McFarland, Frey, & Rhodes, 1980 for further extensions of the generation effect). Notably, we did not find a generation effect in the related group in Experiment 3, where E items were actually recalled better than S items were recalled. One possible interesting source of this reversal in the relative difficulty of E and S items may be related to differences in the nature of the encodings produced in Slamecka and Graf's experiments and our other conditions compared with this one group. The rules used by Slamecka and Graf and those used in our Experiments 1 and 2 may have been less likely to activate as dominant or salient aspects of the meaning of a presented item compared to the instructions given to subjects in the related condition in our Experiment 3. For example, generating "cold" in response to "hot" (which emphasizes the abstract dimension of temperature) may not really activate the most central aspects of the meaning of "hot." That is, "the opposite of cold" may be less central to the concept of hot than "fire" or "burn." While a very specific rule such as "generate an opposite" serves the purpose of equating the nature of the encodings of E and S items, the average encoding probably involves somewhat more peripheral as opposed to more central aspects of the meaning of the words than might otherwise be the case.

In contrast, our related subjects were perhaps more likely to generate responses on the basis of core aspects of the meaning of the stimuli. The production of these responses would not necessarily involve core aspects of the meaning of the responses. If our subjects were more likely to activate core aspects of the stimuli than core aspects of the responses, the E items (the stimuli in Experiment 3) would have an advantage over the S items (the responses) since there is some evidence that memory is better insofar as an encoding involves core features of the meaning of an item (Hasher & Johnson, 1975; Hasher, Griffin, & Johnson, 1977; Hashtroudi, 1977). This line of reasoning suggests one mechanism that might save us from being dominated by our own thoughts in memory. Insofar as internal events are generally of the cotemporal type, *in response* to external events, they will tend to reflect processing for meaning that favors recall of the external event. On the other hand, when E and S items (and core-relatedness of their encodings) are equated (Slamecka & Graf, 1978; Raye et al., 1980), there appears to be a genuine advantage in memory for the subject-generated information.

Reality monitoring

While the present data provide evidence regarding the generation effect, the major purpose of these experiments was to explore aspects of the process of discriminating self-generated from externally derived information in memory. The results of all three experiments are consistent with the proposition that the more closely the responses that are generated are specified by external cues, the harder it is later to discriminate between the externally presented and the internally generated. Thus, in Experiment 1, providing first-letter cues for category instances decreased discrimination relative to no cues. In Experiment 2, low frequency category instances that were presumably a bit harder to generate resulted in better discrimination than high frequency instances. Opposites, which presumably require minimal search and decision processes, resulted in lower discrimination than category instances in Experiment 1, and a smaller difference in confidence ratings given for correct and incorrect source attribution in Experiment 2. In Experiment 3, correct identification of the origin of items was lowest in the condition in which subjects generated responses that were semantically related to the externally presented words, and greatest in the condition in which they generated responses that were not elicited by or related to the presented items. Finally, in Experiments 2 and 3, significant reality monitoring was found after retention intervals of 10 and 7 days, respectively, indicating that there is a remarkable persistence of information upon which origin decisions can be based, even in these laboratory situations in which the homogeneity of the materials should maximize confusion.

Here we have emphasized that one discriminative cue that differentiates the classes of externally presented and self-generated memories is the greater amount of information about cognitive operations typically associated with self-generated traces. Consistent with the hypothesis that such information is used as a discriminative cue is that, across all three experiments, accuracy of reality monitoring was positively related to manipulations designed to increase the information about cognitive operations available for self-generated items. It should also be noted that some manipulations designed to affect the external specification of the response (and thus cognitive operations) might influence other characteristics of memories that, according to our working model, should also affect judgments about the origins of memories. For example, the less an external cue specifies a response, the greater the opportunity for the response to reflect idiosyncratic or personal significance. Reality monitoring judgments should be influenced by considerations of how characteristic of oneself a response is ("Is this the sort of thing I'm likely to say?"). The unrelated condition of Experiment 3 particularly allows for the operation of this factor. The relative contributions of cognitive operations at the time the memory was established, and personal significance judgments at the time of the test for each of the conditions of Experiment 3 cannot be established with the present data. However, we know that idiosyncracy is not the only basis of reality monitoring decisions because significant reality monitoring takes place even when the self-generated target is completely specified by cues (e.g., Raye & Johnson, 1980, Experiment 2; Foley, Johnson, & Raye, Note 2). In fact, a comparison across two experiments that varied greatly in opportunity for personal significance to operate did not suggest that it played any role with materials similar to those used here (Raye & Johnson, 1980). Finally, analyses of the present Experiment 1 that omitted idiosyncratic items did not affect the results; in Experiment 2, a comparison of the identification-of-origin performance for target and idiosyncratic items suggested that the idiosyncratic items were more, not less, likely to be misidentified with respect to origin. This would be expected if idiosyncratic items are responses highly available to the individual and do not require much in the way of cognitive operations to produce. Therefore, pending experimental demonstrations of the role of idiosyncracy in identification of origin decisions in situations such as those used here, the most parsimonious explanation of the relative ordering of conditions in these three experiments is in terms of the relative amounts of information about cognitive operations available for self-generated information.

Further support for this interpretation is that the bias in false positives also can be interpreted with reference to the role of cognitive operations in decision processes. A major finding of the present paper was the marked bias on false positives. Subjects were much more likely to mistakenly identify a new item as externally presented than to mistakenly identify a new item as self-generated. This finding held for three types of items in Experiment 2 and across three instructional conditions in Experiment 3. One source of false positives is presumably information that has been generated during acquisition via mechanisms such as implicit associative responses (Underwood, 1965; Glanzer & Bowles, 1976) or spread of excitation (Collins & Loftus, 1975; Schvaneveldt & Meyer, 1973). If these internal processes produced cognitive operations information, one would expect the items falsely recognized as old to be identified as self-generated. Just the reverse occurs, suggesting that activation via automatic or incidental processes does not produce cognitive operations information of the same sort or to the same degree that activation via voluntary control does. Evidence consistent with this line of reasoning is that explicitly created sensory information (images) was less likely to be attributed to external sources (presented pictures) compared with incidentally created sensory information (Durso & Johnson, 1980). The present false positive data, then, suggest that when an item seems familiar but the trace does not include coded information clearly identifying its source, people have a bias to conclude that it was external in origin. People seem to assume that they will remember what they themselves have generated (as indicated by subjects' predictions in Experiment 2 and the higher confidence ratings associated with S-responses in Table 4), and, hence, if they do not remember producing a familiar thought, conclude it must have originated elsewhere.

Finally, with respect to reality monitoring, the subject-report data were strikingly consistent with the general assumptions of the model. Although these comments should be followed up more formally within a structured questionnaire, there was a clear tendency for sensory information to be spontaneously mentioned more frequently in connection with external events and for reference to cognitive operations to be more frequent with reference to internally generated information. In addition, there were some comments which seemed to describe both the attribution bias reflected in the false positive data and the reasoning processes postulated by the model.

Notes

Requests for reprints may be sent to Marcia K. Johnson, Department of Psychology, State University of New York, Stony Brook, NY 11794. Received for publication April 28, 1980; revision received August 8, 1980. This research has been supported by a grant from the National Science Foundation.

1. All statistical tests of significance are reported at the .05 level.

Reference notes

1. Johnson, M. K., & Raye, C. L. A working model of reality monitoring. Annual Meeting of the Psychonomic Society, November, 1978.
2. Foley, M. A., Johnson, M. K., & Raye, C. L. A developmental study of reality monitoring, comparing two types of judgments about memories. Annual Meeting of the Eastern Psychological Association, April, 1980.

References

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, *80*, (3, Pt. 2.)

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, *2*, 331–350.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, *82*, 407–428.

Durso, F. T., & Johnson, M. K. The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 416–429.

Glanzer, M., & Bowles, N. Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 21–31.

Hasher, L., Griffin, M., & Johnson, M. K. More on interpretive factors in forgetting. *Memory & Cognition*, 1977, *5*, 41–45.

Hasher, L., & Johnson, M. K. Interpretive factors in forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *1*, 567–575.

Hashtroudi, S. Analytic versus synthetic encoding and long-term retention. (Doctoral dissertation, State University of New York, Stony Brook, 1977). *Dissertation Abstracts International*, 1977, *38*, 1921B–1922B. (University Microfilms No. 77-20,953.)

Jacoby, L. L. On interpreting the effects of repetition: solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 1978, *17*, 649–667.

Johnson, M. K. What is being counted none the less? In I. M. Birnbaum & E. S. Parker (Eds.), *Alcohol and human memory*. Hillsdale, N.J.: Erlbaum, 1977.

Johnson, M. K., & Raye, C. L. Reality monitoring. *Psychological Review*, in press.

Johnson, M. K., Raye, C. L., & Durso, F. T. Reality monitoring: Second perceptions and thoughts. *Bulletin of the Psychonomic Society*, 1980, *15*, 402–404.

Johnson, M. K., Taylor, T. H., & Raye, C. L. Fact and fantasy: The effects of internally generated events on the apparent frequency of externally generated events. *Memory & Cognition*, 1977, *5*, 116–122.

Loftus, E. F. Leading questions and the eyewitness report. *Cognitive Psychology*, 1975, *7*, 560–572.

McFarland, C. E., Frey, T. J., & Rhodes, D. Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 210–225.

Raye, C. L. Recognition: Frequency or organization? *American Journal of Psychology*, 1976, *89*, 645–658.

Raye, C. L., & Johnson, M. K. Reality monitoring vs. discriminating between external sources of memories. *Bulletin of the Psychonomic Society*, 1980, *15*, 405–408.

Raye, C. L., Johnson, M. K., & Taylor, T. H. Is there something special about memory for internally generated information? *Memory & Cognition*, 1980, *8*, 141–148.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, *2*, 437–442.

Schvaneveldt, R. W., & Meyer, D. E. Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), *Attention and Performance IV*. New York: Academic Press, 1973.

Slamecka, N. J., & Graf, P. The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 592–604.

Underwood, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, *70*, 122–129.