Journal of Experimental Psychology: General
1989, Vol. 118, No. 3, 255–257

0096-3445/89/$00.75

How Much Does Guessing Influence Recall? Comment on Erdelyi, Finks, and Feigin-Pfau

Henry L. Roediger III and Kavitha Srinivas
Rice University

Paula Waddill
Purdue University

Erdelyi, Finks, and Feigin-Pfau (1989) present evidence that variations in recall criteria can affect the number of items correctly recalled. In this comment, we (a) describe some procedural differences between their work and the earlier experiments of Roediger and Payne (1985), (b) note that their large manipulations of recall criteria produced only small effects on the amount recalled, and (c) describe recent research complementing that of Erdelyi et al. We observe that variations in recall criteria have larger effects after a 1-week delay than on an immediate test.

A common assumption among researchers investigating memory is that variations in response criteria (manifested in intrusion rates) affect the amount of material recalled. So, for example, if subjects recall more material under some condition (e.g., hypnosis) than under a control condition, but the intrusion rate is also greater under the experimental condition, then the greater recall is likely to be attributed to guessing (e.g., Dywan & Bowers, 1983; Klatzky & Erdelyi, 1985).

The assumption that recall increases as subjects relax their criteria for producing responses is intuitive, but remarkably little evidence for it exists. Roediger and Payne (1985) examined this assumption by presenting subjects with a mixed list of words varying in concreteness and testing them under three conditions: (a) free recall, with a warning against guessing; (b) "uninhibited" recall, in which they were to recall the list but to free-associate and to guess while doing so; and (c) forced recall, in which they were told to recall the list but to produce a large fixed number of responses, guessing if necessary. This manipulation produced a large variation in intrusion rates, but it had no effect on the number of items correctly recalled. Erdelyi, Finks, and Feigin-Pfau (1989, Experiments 1 and 2) replicated these findings with Roediger and Payne's (1985) materials.

In other experiments, Erdelyi et al. (1989, Experiments 3 and 4) report the first systematic evidence that variations in recall criteria can affect levels of recall. The purpose of this commentary is to consider (a) some methodological issues and some differences between Erdelyi et al.'s procedures and those of Roediger and Payne (1985), (b) the magnitude of the effects and their importance for conducting memory research, and (c) some new data on these issues.

The materials and procedures in Erdelyi et al.'s (1989) Experiments 3 and 4 differed in several ways from those of Roediger and Payne (1985). First, Erdelyi et al. used pictures or concrete words rather than more abstract words. Second, they used guessing controls in which subjects who had not seen the original list were yoked to the experimental subjects by being given the experimental subjects' recall protocols and asked to generate further responses. This guessing control seems appropriate and should prove useful in future studies in which the problem of variations in recall criteria may be significant. A third difference is that Roediger and Payne provided forced-recall instructions (or, to another group, the "uninhibited" recall instructions) at the beginning of the test, whereas Erdelyi et al. provided forced-recall instructions only after a period of free recall. Erdelyi et al.'s procedure is methodologically cleaner and avoids any processing bias of the sort that they describe, but it may not faithfully mimic conditions under which response biases are likely to operate; that is, when subjects are tested under hypnosis, under the influence of drugs, or at great delays when memory is likely to be error prone, any response bias is likely to be induced at the beginning of the test and to operate throughout it, rather than to be confined to a period at the end of the test. Thus it may still be of interest to examine the effect of variations in recall criteria when these are instantiated at the beginning of the test rather than in the middle.

If an investigator is worried that variations in recall criteria may undermine an experimental investigation of memory, Erdelyi et al.'s (1989) experiments point to one sure way to avoid the problem: Use study materials that subjects cannot guess. This was achieved by Roediger and Payne (1985) in their list of high- and low-imagery words, because Erdelyi et al. showed (by their control procedure in Experiments 2A and 4A) that there was a low probability that subjects would guess these items. Thus if one is interested in memory processes uncontaminated by guessing, use of such materials will avoid the problem altogether, even when conditions differ in their guessing base rates. Of course, the applicability of this solution will depend on the purposes of the investigation. If researchers are interested in whether eyewitnesses to crime can recall more under the influence of hypnosis than without it, then medium-frequency, low-imagery words would not be the material of choice. However, for purposes of investigating many basic issues in the laboratory, materials that are hard to guess would be appropriate in order to avoid the influence of variations in the recall criteria.

Correspondence concerning this article should be addressed to Henry L. Roediger III, Department of Psychology, Rice University, P.O. Box 1982, Houston, Texas 77251-1892.

In the prior paragraphs we treated Erdelyi et al.'s (1989) results as if they showed that variations in recall criteria could produce significant problems in interpreting recall results. However, in some sense this remains to be demonstrated. Although Erdelyi et al. did show that variations in recall criteria could affect the amount recalled, their large manipulations of criteria produced only small gains in correct recall. Presented in Table 1 is a summary of their primary results showing the effects of recall criteria in Experiments 3 and 4. (The mean values at the bottom of the table weight each condition equally). The mean difference between forced and free recall in intrusion rates was 26.51 (i.e., 26.51 more intrusions in forced than in free recall). The corresponding difference in correct recall (R_b) was 2.61. The ratio of these values (R_b/intrusions) was about .10. Thus, subjects in the forced-recall condition correctly recalled one item for every 10 guesses or intrusions.

To cast the argument in terms of signal detection theory, albeit very loosely, the false-alarm rates between forced- and free-recall conditions differed by a factor of 25 (27.58 to 1.07) and yet the hit rate was influenced by a factor of only 1.65 (6.58 to 3.97). It is difficult to think of any other area of investigation in classical psychophysics or in recognition memory in which so large an effect on false alarm rates has resulted in so small an effect on hit rates. Were memory operating characteristic functions drawn for the data, they would be lines that had only the shallowest slope. Thus even with Erdelyi et al.'s (1989) positive results, the contamination of correct recall by chance hits resulting from guessing may not be too great, at least with these types of material. Corrections for guessing in recall often have little impact on the conclusions drawn (e.g., Tulving & Pearlstone, 1966).

We raise two other questions about the effects shown by Erdelyi et al. (1989). These questions are how much forced recall actually aids memory once "pure guessing" is eliminated and whether variations in recall criteria would have greater effects on delayed tests than on immediate tests. Consider the guessing issue first. Even in Experiments 3 and 4, in which Erdelyi et al. found an advantage of forced over free recall, they did not find any benefit in relation to the expected guessing rates. For example, the advantages of forced over free recall were 2.61 items for pictures and 2.89 items for words. However, the advantages expected on the basis of pure guessing from the control groups in Experiment 4A were 5.17 and 3.22 items for pictures and words, respectively. Unfortunately, we cannot know simply from the advantage in number correct in forced recall whether the subjects are remembering more items than expected by chance. Because Erdelyi et al.'s subjects were recalling either as many items (words) as or fewer items (pictures) than expected from the controls, the additional hits may not have represented true memories.

Recent experiments by Roediger and Challis (1989, undated) have addressed these issues more directly. Here we present illustrative data from selected conditions of one experiment. In the relevant conditions, four groups of subjects were presented with 60 pictures and then tested under free-recall instructions (two groups) or forced-recall instructions (the other two groups). Two groups given each type of test were tested both soon after presentation and after a 1-week delay, whereas the other groups were tested only after the 1-week delay. Subjects were given 10 min for their tests, and forced-recall subjects were required to write down 60 items and were encouraged to produce even more if they could. Free-recall subjects were warned not to guess but were asked to continue trying during the entire test period. After each test was completed, subjects in all groups were asked to give confidence ratings indicating the likelihood that each produced item had actually occurred in the study list. Items were rated on a 6-point scale (6 = *high confidence that responses had appeared on the study list* and 1 = *high confidence that responses had not appeared on the study list*).

The relevant results are presented in Table 2. As shown in the top two rows, forced-recall subjects produced many more intrusions (35.6) than did free-recall subjects (1.8) on the immediate test. In addition, forced-recall subjects recalled more correct items (32.8) than did free-recall subjects (29.2), which shows the advantage of forced over free recall with pictorial materials, as Erdelyi et al. (1989) also showed. However, a further analysis of the correct responses showed that when conditionalized on subjects' recognition that the response was indeed an old item (ratings of 4, 5, and 6 on the scale), the advantage of forced over free recall disappeared (28.8 correct vs. 28.2). Thus it seems that even when forced-recall subjects produce more correct responses than do free-recall subjects, they may not remember any more information. Instead, the few additional "correct" responses may reflect free-association priming, as in studies of implicit memory (see Schacter, 1987). Forced-recall subjects may produce extra studied items, but they do not realize that these are from the studied list. Exactly the same pattern of results occurred for these subjects 1 week later. Now the advantage of forced over free recall in producing correct items was even greater (29.2 to 23.2), but when conditionalized on the basis of confidence ratings, the effect disappeared (21.0 to 20.9).

One implication of these results is that although the forced-recall procedure may indeed cause subjects to lower their response criteria, it also causes subjects to confabulate and

Table 1
Selected Data From Erdelyi, Finks, and Feigin-Pfau's (1989) Experiments 3 and 4

Experiment	Material	R_b: Forced recall	R_b: Free recall	R_b: Difference	Intrusions: Forced recall	Intrusions: Free recall	Intrusions: Difference	Ratio of differences: R_b/intrusions
3	Pictures	6.53	4.18	2.35	25.18	0.65	24.53	0.096
3	Pictures	7.22	4.61	2.61	25.28	1.33	23.95	0.109
4	Words	6.00	3.11	2.89	32.28	1.22	31.06	0.093
M		6.58	3.97	2.61	27.58	1.07	26.51	0.098

Table 2
Results From Roediger and Challis (undated)

	Time of test					
	Immediate			1-week delay		
Recall condition	Correct	Judged "old"[a]	Errors[b]	Correct	Judged "old"[a]	Errors[b]
Free	29.2	28.2	1.8	23.2	20.9	3.1
Forced	32.8	28.8	35.6	29.2	21.0	48.2
	Delayed test only					
Free				16.7	11.5	5.3
Forced				25.9	15.1	42.5

[a] The number of correctly recalled items that subjects later judged to have been from the study list (ratings of 4, 5, or 6 on a scale from 6 = *high confidence that responses had appeared on the study list* to 1 = *high confidence that responses had not appeared on the study list*).
[b] The number of intrusions.

consequently to be confused on later tests as to whether produced items were studied events or were generated to fill up the page on the earlier test. This failure of reality monitoring (Johnson & Raye, 1981) can be seen clearly when free-recall performance is compared with forced-recall performance on Test 2. Free-recall subjects, who produced very few intrusions on the early test, correctly recognized 90% of their correct responses on Test 2 as having been on the list (20.9 of 23.2). On the other hand, forced-recall subjects recognized only 72% of their correct responses as having been on the list during Test 2. Although forced-recall procedures enable one to control for response biases, they may also undermine subjects' memories for actual events by requiring them to supply interference in the form of generated responses to satisfy the experimenter's demand.

One pattern of results from the two groups tested after a 1-week delay is quite different (see Table 2). Forced-recall subjects again produced many more intrusions and more correct responses than did free-recall subjects, but now the advantage of forced over free recall survived conditionalization on the confidence ratings (15.1 to 11.5 correct). We have confidence in this pattern both because it was statistically significant and because we have replicated it under slightly different study conditions. It may well be that the problem of variations in recall criteria are greater after a delay than on an immediate test (or perhaps greater under any set of circumstances that produces poor retention). The reasons for this state of affairs are not obvious, but retention interval provides at least one variable that may allow researchers to control criteria effects and may lead to further discoveries.

References

Dywan, J., & Bowers, K. (1983). The use of hypnosis to enhance recall. *Science, 222*, 184–185.

Erdelyi, M. H., Finks, J., & Feigin-Pfau, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General, 118*, 245–254.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 57–85.

Klatzky, R. L., & Erdelyi, M. H. (1985). The response criterion problem in tests of hypnosis and memory. *International Journal of Clinical and Experimental Hypnosis, 33*, 246–257.

Roediger, H. L., III, & Challis, B. H. (1989). Hypermnesia: Improvements in recall with repeated testing. In C. Iwaza (Ed.), *Current issues in cognitive processes: The Tulane-Floweree symposium on cognition* (pp. 175–199). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, & Challis, B. H. (undated). *Varying recall criteria has differential effects on immediate and delayed recall.* Manuscript in preparation.

Roediger, H. L., III, & Payne, D. G. (1985). Response criteria do not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition, 13*, 1–7.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381–391.

Received October 17, 1988
Revision received December 30, 1988
Accepted January 12, 1989 ■