Recognizing Differences in Recognition Tasks:  
A Reply to Lassiter and Slaw

Catherine Hanson  
Temple University

William Hirst  
New School for Social Research

Hanson and Hirst (1989) recently provided evidence that parsing strategies affect Ss' ability to recall but not to recognize episodic information. This finding conflicts with results obtained by Lassiter (1988) and Lassiter, Stone, and Rogers (1988) as noted by Lassiter and Slaw (1991). On the basis of a meta-analysis combining data from the Hanson and Hirst and the Lassiter experiments, Lassiter and Slaw argued that the discrepancy in findings across studies is "more illusory than real" (p. 81). In this article, four critical differences between the Hanson and Hirst study and the two Lassiter studies are outlined. It is argued that these factors must be considered when drawing conclusions about the effect of parsing strategies on recognition performance.

Memory research over the past century has concerned itself with two fundamental questions. First, how is personal experience recorded in the mind? Second, by what means is recorded information accessed? In information processing terms, memory research is concerned with encoding and retrieval processes. Over the years, considerable effort has been spent not only in understanding how information is encoded and retrieved, but also in examining how the two processes may interact.

One thing researchers have learned about retrieval processes is that the way memory is probed will greatly affect what can be accessed. For example, it is well known that recognition tasks typically yield better memory performance than do recall tasks (see Brown, 1976, for a discussion of this issue). However, under the right conditions, recognition performance can be depressed. The similarity of foils to targets (Goldstein & Chance, 1970), the semantic relatedness of targets and foils (Underwood, 1965), and the number of distractors (Davis, Sutherland, & Judd, 1961) are but a few of the factors that have been shown to influence recognition performance.

We recently used a parsing task to affect the way subjects orient toward a videotaped event sequence (Hanson & Hirst, 1989). Subjects were asked to indicate the end of one event and the beginning of another, with subjects in one condition oriented toward large events and subjects in another condition oriented toward small events. Recall and recognition of the episode was then assessed. Our results indicated that the level of orientation affected recall performance, but had no notable effect on recognition.

Lassiter and Slaw (1991) argued that the difference between recognition and recall performance observed by Hanson and Hirst (1989) in their study of event parsing was probably an artifact of sample size. They based their position on comparisons between studies conducted by Lassiter (1988) and Lassiter, Stone, and Rogers (1988) and by Hanson and Hirst.

Whereas the Lassiter studies found that fine unit parsers both recognize and recall more of an event sequence than gross unit parsers, Hanson and Hirst found that fine unit parsers recall more of the event sequence but do not recognize more than gross unit parsers. Lassiter and Slaw's basic argument is that had Hanson and Hirst used a larger sample size, they would have found recognition differences between fine unit and gross unit parsers as had the Lassiter studies, which used larger numbers of subjects.

We do not find the Lassiter and Slaw (1991) argument convincing for a number of reasons. These researchers base their position on a meta-analysis involving three Lassiter experiments (Lassiter, 1988, Experiment 1; Lassiter et al., 1988, Experiments 1 & 2) and three of the Hanson and Hirst experiments (1989; Experiments 1, 2, & 3). The underlying assumption in this type of analysis is that all experiments under consideration are essentially alike. That is, the fine and not so fine differences among the experiments being analyzed can be ignored. As noted earlier, however, it is well known by memory researchers that what may appear to be minor changes in methodology can greatly affect performance in a recognition task. Although Lassiter and Slaw seem to assume that only sample size distinguishes the Lassiter experiments from our own, we note four differences between the Lassiter studies and the Hanson and Hirst studies that we feel contribute to the discrepancy between Lassiter's findings and our own.

First, whereas the three Hanson and Hirst experiments were specifically designed to compare recall and recognition performance and used a within-subject design for this purpose, a comparison across recall and recognition was not the primary goal of the Lassiter (1988) and Lassiter et al. (1988) experiments and is possible only by comparing across subjects and experiments. Our assumption in using a within-subject design was that the effect size for recognition differences should be comparable to that for recall differences. Thus, if our sample size was large enough to observe a difference between fine unit and gross unit parsers in one memory measure, it should be sufficient to observe any difference in the other memory measure. By arguing that we did not use a large enough sample size, Lassiter and Slaw (1991) are implicitly suggesting

Correspondence concerning this article should be addressed to Catherine Hanson, Weiss Hall, Psychology Department, Temple University, Philadelphia, Pennsylvania 19122. Electronic mail may be sent to cat@euclid.math.temple.edu.
that the effect size for recognition differences should be smaller than that for recall differences. They provide no basis for this assumption.

Second, subjects in the Lassiter experiments (Lassiter, 1988, Experiment 1; Lassiter et al., 1988, Experiments 1 & 2) performed neither as well nor as consistently as those in the Hanson and Hirst (1989) experiments. Over the three Hanson and Hirst recognition tasks, subjects reliably recognized an average of 73%--77% of items, averaging 75% overall. In Experiment 1 of the Lassiter study, recognition performance ranged from a low of 51% to a high of 62%, averaging 57%. In Experiment 1 of the Lassiter et al. study, subjects' recognition performance ranged from 56% to 65% across conditions, averaging 58%. More significantly, in Experiment 2, correct recognition ranged from a low of 48% to a high of 88%, averaging 68% across conditions. Interestingly, fine unit parsers were responsible for scores at both extremes in this experiment.

Third, whereas Hanson and Hirst provide some detail about the selection of items for their recognition task (e.g., both true and false items were rated as typical of the event sequence by an independent group of subjects before the experiments proper), Lassiter (1988) and Lassiter et al. (1988) provide neither a rationale for item selection, nor any information about the construction and composition of questionnaires beyond a cursory example item. Thus, it is not clear that the Lassiter and Lassiter et al. tasks were probing the same kind of information at the same level of detail that the Hanson and Hirst task was. The impact of target and distractor selection on recognition performance can be seen clearly in Experiment 2 of Lassiter et al. (1988). In this experiment, five items of the recognition task were aimed at probing “action-related” memory, and five were aimed at probing “nonaction-related” memory. Fine unit parsers recognized significantly more action items than gross unit parsers but significantly fewer nonaction items. Collapsing over the action–nonaction factor and a second factor concerned with the effect of leading questions, we find the average recognition performance by fine unit parsers (68%) to be similar to that of gross unit parsers (69%). Clearly, any conclusions that can be drawn concerning the recognition performance of fine unit and gross unit parsers must take into account the nature of to-be-remembered material and test distractors.

Finally, and most importantly, the format of the recognition tasks used by Lassiter (1988), Lassiter et al. (1988), and Hanson and Hirst (1989) varied widely. In the three Hanson and Hirst experiments, the same 24 true–false item questionnaire was used. In Experiment 1 of Lassiter et al., a 9-item, four-choice recognition task was used, and in Experiment 2 of that same study, a 10-item, yes–no recognition task was used. Experiment 1 of the Lassiter study used a 27-item, five-choice recognition task. The effect of parsing on recognition that Lassiter and his colleagues found may be a function of their use of multiple choice recognition tasks. It is known that recall and recognition become more difficult to differentiate with increasing choices in a recognition task (Brown, 1976), and interestingly, the experiments in the cited Lassiter work that yield differences between fine unit and gross unit parsers were those in which multiple choice recognition tasks were used. The one cited experiment in which a yes–no recognition task was used (Lassiter et al., 1988, Experiment 2) did not yield consistent differences in recognition performance between fine unit and gross unit parsers.

As we noted in our General Discussion (Hanson & Hirst, 1989), it is possible that our recognition task was not sufficiently sensitive to pick up differences between fine unit and gross unit parsers and that further investigation of this issue is warranted. Clearly, if memory researchers are to gain any general understanding of how and what kind of episodic information is remembered, they must adopt a systematic, analytic approach to test construction and item selection, drawing on what is already known about encoding and retrieval processes to guide them.

References


Received October 15, 1990
Accepted November 30, 1990