

# Reminiscence and arousal: Replications and the matter of establishing a phenomenon

WILLIAM H. SAUFLEY, JR.

*Institute of Human Learning, University of California, Berkeley, California 94720*

and

SUSAN C. LaCAVA

*Emory University, Atlanta, Georgia 30322*

The experiment presented here was designed to produce reminiscence of "high-arousal" (high-GSR) verbal material but failed to do so. An outline for classifying replications and their role in establishing a phenomenon was used to evaluate this experiment and several related ones. It was concluded that the arousal-recall effect contains more than one phenomenon and is dependent upon unique experimental conditions not yet specified.

Several years ago some experiments produced changes in recall of as much as 400%. Under any circumstances an effect of that magnitude demands interest, but doubly so when the change takes place without the active imposition of an experimental treatment and, moreover, is an increase in recall with passing time—reminiscence. Here we present an effort to produce the effect and discuss the problems of interpreting replications.

The experimental search derives from a theoretical interpretation, termed "action decrement," advanced by Walker (1958) during the early period of research on generalized cortical arousal and learning. When applied to memory for verbal material, a differential effect is predicted for retention, namely a low probability of immediate recall and a high probability of delayed recall of high-arousal material, with the converse for low-arousal material. Kleinsmith and Kaplan provided one of the first tests of the hypothesis in 1963. The task was one-trial learning of eight word-digit pairs under incidental learning conditions. A slow presentation rate was used to permit separation of the arousal effects and measurement of the GSR to each pair for classification (within subjects) of one of two categories: high arousal and low arousal. Recall tests, taken from 2 min to 1 week following learning, yielded the predicted interaction between arousal level and recall over time. Moreover, there was an absolute increase in recall for the high-arousal items.

This experiment established a phenomenon, the

This research was carried out at Emory University and the paper prepared by the senior author while a visitor at the Institute of Human Learning, University of California, Berkeley. The paper was materially improved by suggestions from Geoffrey Keppel who, although blameless for the contents, agreed to accept editorial responsibility.

interaction of high and low arousal with recall over time. The interaction had two outstanding features: an absolute increase in recall probability with passing time for the high-arousal items and no main effect of forgetting over the 1-week interval. In other words, the interaction was balanced so that when the within-subject variable, high and low GSR, was disregarded, recall probability remained statistically constant. One experiment followed the Kleinsmith and Kaplan (1963) procedures almost exactly (Butter, 1970, Experiment 1), used the same stimulus items, and found the same results: an interaction and no main effect of delayed testing (i.e., forgetting) over 2 days. Other studies (Kaplan & Kaplan, 1969; Kleinsmith & Kaplan, 1964; Walker & Tarte, 1963) have produced generally similar results but not always an absolute increase in recall for the high-arousal items. However, instructions, normative or ipsative scoring, within- or between-subject designs (equivalent to mixed or unmixed lists) are procedural variations which could have acted alone or in combination to produce the various pattern of results. Such changes in results with changes in procedure, while offering potential generalization, render the time course of the reminiscence difficult to specify. Recently, three studies (Deffenbacher, Platt, & Williams, 1974; Geen, 1973, 1974) have tested the theory by interpreting arousal as "social arousal" and found reminiscence along with no mean forgetting. Social arousal was imposed in a between-groups design, using the same stimulus material as Kleinsmith and Kaplan (1964), with a much shorter presentation rate, no interpair interval, and intentional learning instructions. Other studies have reported results consistent or inconsistent with the arousal-recall interaction but have great procedural differences from those cited here.

We are concerned with the experimental conditions used by Kleinsmith and Kaplan through which the

arousal-recall effect was introduced. Our original object was to produce the effect then, subsequently, to determine the rate of reminiscence as a function of time and experimental variables. We made two preliminary attempts which followed Kleinsmith and Kaplan (1964) but which failed to produce the arousal-recall effect. Two difficulties appeared. First, recall was consistently low, with mean correct responses varying from .25 to .90. Second, a basal resistance-level deflection occurred during the interpair interval. The 8-sec interpair interval was filled with color naming to allow the GSR to stabilize and to reduce chances of a carryover from one pair to the next. The problem was persistence of the GSR from the color-naming activity to the next CVC-digit pair.

Two changes were made to raise the level of performance: Subjects were instructed to learn and different CVCs were selected to reduce the formal similarity in the stimulus items. The six taken from Kleinsmith and Kaplan used 13 out of 18 different letters while the six used here had 17 out of 18 different letters. Also, an experimental treatment was included in which the interpair interval did not involve color naming.

## METHOD

### Subjects

The subjects were 96 undergraduates obtained from the introductory psychology class. They were assigned to the groups in random blocks of eight as they appeared for testing, resulting in eight groups of 12 subjects each.

### Apparatus

The GSR was measured by means of an E&M Instrument physiograph. A 1-cm-square hole cut in masking tape was placed over the middle phalange of the freshly washed index and ring fingers of the nonpreferred hand. The paste (Edelberg & Burch, 1962) was applied and electrodes were placed over the holes, secured with masking tape, and wrapped with two-ply surgical gauze to provide heat insulation. Metal bird-cage splints were placed over this to encourage the subject to hold the fingers still. The response was recorded at a chart paper speed of .2 cm/sec. The recording pen was calibrated by momentarily inserting a positive 2,500-ohm resistance in series with the subject and adjusting for a 3-cm pen deflection.

### Lists

The following list of 22% association value trigrams was chosen from Archer (1960), with formal similarity kept low: TYB, FIH, OOC, ZEF, PUV, DAJ. The response items were the single-digit numbers from two to seven. Four different sequences of the pairs were used and recall orders were varied by giving each subject one of the orders different from his own.

### Procedure

Stimulus material was presented on slides at a 4-sec rate. A single learning trial was given, with each stimulus term presented alone followed by the stimulus-response pair. There was an 8-sec interval between the pairs in order to separate arousal effects of one pair from the next. Two interpair activities were used: a color-naming task or no intervening task. During the color-naming period, two slides, each containing four color squares, were shown and subjects were asked to name the colors. If there was no intervening task, the subjects were simply asked to look at the dark screen. All subjects were instructed to learn the trigram-number associations but were urged not to rehearse

during the 8-sec interpair interval and to try and think of something else if they thought of the items during the interval.

Two recall conditions, paced and unpaced, were used. Paced recall consisted of presenting the stimuli one at a time for 4-sec intervals and having the subject guess the number that went with it. Since the response terms were not presented in the test phase, a 12-sec interstimulus interval was used. During this interval the subject either named colors or the screen was dark, whichever was appropriate for the condition. The unpaced recall test consisted of presenting a list of the trigrams on a page with spaces for writing the appropriate digit next to the trigram. Twenty-four seconds were allowed to complete recall. Two recall intervals were used: 2 min and 1 week. Guessing was encouraged and the electrodes were in place during recall.

## RESULTS

The scoring interval used to determine the arousal value of each pair was 12 sec and began with presentation of trigrams alone (Kaplan & Kaplan, 1968). Each of the six GSR deflections was ranked. For each subject the three largest deflections were designated "high-arousal learning" and the three lowest as "low-arousal learning." Any ties were broken randomly.

Reminiscence did not appear. For the high-arousal items there were 1.50 and .77 mean correct, for low arousal 1.42 and .75, at the 2-min and 1-week tests, respectively (each of these means is based upon 48 subjects). The arousal-recall interaction was not significant ( $F < 1$ ); in fact, no interaction was significant. To make certain that the method of separating items into "high" and "low" categories did indeed produce reliably different groups, the GSR scores were converted to logs. The pairs classified high arousal had reliably greater scores than those classified low arousal [ $t(95) = 11.7, p < .001$ ]. Since data from our previous experiment indicated the presence of a GSR to the color-naming task, a separate analysis of the color-naming group in this experiment was made to find out if the classification was reliable. The difference between high- and low-arousal pairs was still reliable [ $t(46) = 14.38, p < .01$ ].

There was significant forgetting [ $F(1,88) = 14.38, p < .01$ ], and the color-naming condition produced fewer correct responses than the dark-screen condition [ $F(1,88) = 5.61, p < .05$ ]; neither of these is of particular importance since none of the interactions was significant. Both recall and GSR scores were scored with respect to serial position during the study trial. Recall scores did not exhibit systematic effects but the greatest GSR uniformly occurred to the first pair (also reported by Walker & Tarte, 1963).

## DISCUSSION

Significant forgetting was found over the 1-week retention interval, a typical result from memory research. There was a failure to find the arousal-recall interaction in spite of our efforts to establish what we thought to be the essential aspects of the experimental conditions and the use of a large number of subjects for the critical comparisons. In order to investigate a phenomenon such as reminiscence, it is necessary to describe it in terms of orderly response variations over time, but it is not

possible to make such a description from the data currently available. The issue with which we are confronted is one of determining when a phenomenon is established, of determining when an effect is dependable and capable of being generalized, mitigated only by theoretically relevant variables. This work begins with the original experiment and continues through each attempted replication. To judge the effects of a replication, it is necessary to consider both the form of replication and the status of the effect under investigation.

At least three categories of replications can be identified, each possessing a trade-off between bringing out new information and confirmation or lack of confirmation of the original effect. *Exact* replications duplicate in detail the original experimental procedure, and thus offer the greatest reason for doubt when failing to produce the effect but little new information when reproducing a positive result. Exact replications require careful communication with the original investigator, an additional reason exact replications are seldom done. *Partial* replications, including the present experiment, introduce enough procedural differences from the original to reduce the impact of a failure but to be valuable for expanding and analyzing a positive effect.<sup>1</sup> *Generic* replications manipulate the theoretically active variables by using different experimental manipulations. This class of experimentation makes the greatest contribution toward testing the theory and generalizing the original phenomenon but is of small consequence if it fails to produce the sought-after effect. Butter's Experiment 1 (1970) is the closest to an exact replication of Kleinsmith and Kaplan (1963) and confirmed the original effect. The other experiments using the GSR are partial replications. Those manipulating social arousal can be classified as generic replications.

Implications of a failure to produce an effect depend upon the status of the effect investigated. If the effect is an established one, widely tested and frequently produced, a failure could point to unique circumstances which prevent the standard result from occurring. On the other hand, if the effect is not well established, a failure could signal that singular conditions may be required to tease out the phenomenon.

Because the effect has been found with a growing variety of experimental procedures, it appears to be robust. However, in part this could be a result of self-selection of reports by researchers and editorial tendency to publish positive results. Although such biases apply to all research topics, their importance looms before a novel phenomenon. Next, an examination of the results of available experiments indicates more than one phenomenon might be present, reminiscence or differential forgetting. Moreover, the necessary experimental technique is elusive. In order to evaluate our results, a brief survey of the GSR experiments is necessary. The Walker and Tarte (1963) experiment is excluded here because the method of data presentation makes it not comparable. The two original experiments (Kleinsmith & Kaplan, 1963, 1964) and Experiment 1 of Butter (1970) found reminiscence and no forgetting. On the other hand, Butter's Experiment 2, using different stimuli, found reminiscence (a finding not replicated by Schmitt and Forrester, 1973) with slight but significant forgetting, while Kaplan and Kaplan (1969) found differential forgetting with a variation in instructions. With the exception of Butter's Experiment 2, the stimuli have been either the same set of CVCs or words. Finally, the three social arousal experiments found reminiscence, no forgetting, and did use the original CVC stimuli.

Changes we made in procedure in an effort to improve the technique may have been sufficient to puncture the effect. We used different CVCs as stimuli, intentional learning instructions, and a different method of scoring the GSR. (Our two preliminary experiments even without these modifications did not show the effect.) Evidently, two exceptional results occur together: reminiscence and no forgetting for the list as a whole.

Each is rare and signals a unique set of experimental conditions. A reference to several unpublished studies by Kaplan and Kaplan (1969) would seem to bear this out. Therefore, certain delicate experimental conditions may be necessary to produce reminiscence. For example, given the number of experiments, a restricted range of stimuli has been used: either the same set of words or CVCs. Instructions have also varied. We fully instructed subjects and attempted to set them at ease for the experimental session. The studies employing social arousal, as well as Kaplan and Kaplan (1969), indicate that instructions and experimental conditions which induce stress may bear upon the subjects' perception of the task so as to constitute a critical factor.

No experimental finding rests upon one or two experiments but upon data gathered under a variety of conditions. A failure, especially a partial replication, does not counterweigh a positive result even when Type II error is ignored. Nevertheless, a failure does serve notice that the sought-after phenomenon, say, reminiscence, is dependent upon circumscribed conditions perhaps not clearly understood. It also can cause other investigators to reexamine replications and perhaps pinpoint the effective variables. Such a need is buttressed when a "normal" result (forgetting) is found in replication but not the novel and counterintuitive results (no forgetting, reminiscence) of the original reports.

#### REFERENCES

- ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, 74(Whole No. 10).
- BUTTER, M. J. Differential recall of paired-associates as a function of arousal and concreteness-imagery levels. *Journal of Experimental Psychology*, 1970, 84, 252-256.
- DEFFENBACHER, K. A., PLATT, G. J., & WILLIAMS, M. A. Differential recall as a function of socially induced arousal and retention interval. *Journal of Experimental Psychology*, 1974, 103, 809-811.
- DINNER, J. E., & DUNCAN, C. P. Warm-up in retention as a function of degree of verbal learning. *Journal of Experimental Psychology*, 1959, 57, 257-261.
- EDELBERG, R., & BURCH, N. R. Skin resistance and galvanic skin response. *Archives of General Psychiatry*, 1962, 7, 25-31.
- GEEN, R. G. The effects of being observed on short- and long-term recall. *Journal of Experimental Psychology*, 1973, 100, 395-398.
- GEEN, R. G. Effects of evaluation apprehension on memory over intervals of varying length. *Journal of Experimental Psychology*, 1974, 102, 908-910.
- IRION, A. L. Retention and warming-up effects in paired associate learning. *Journal of Experimental Psychology*, 1949, 39, 669-675.
- KAPLAN, R., & KAPLAN, S. The arousal-retention interval interaction revisited: The effects of some procedural changes. *Psychonomic Science*, 1969, 15, 84-85.
- KAPLAN, S., & KAPLAN, R. Arousal and memory: A comment. *Psychonomic Science*, 1968, 12, 73-74.
- KLEINSMITH, L. J., & KAPLAN, S. Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*, 1963, 65, 190-193.
- KLEINSMITH, L. J., & KAPLAN, S. Interaction of arousal and recall interval in nonsense syllable paired-associate learning. *Journal of Experimental Psychology*, 1964, 67, 124-126.
- LAZAR, G. Warm-up before recall of paired adjectives. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 321-327.

- ROCKWAY, M. R., & DUNCAN, C. P. Pre-recall warming up in verbal retention. *Journal of Experimental Psychology*, 1952, **43**, 305-312.
- SCHMITT, J. C., & FORRESTER, W. E. Effects of stimulus concreteness-imagery and arousal on immediate and delayed recall. *Bulletin of the Psychonomic Society*, 1973, **2**, 25-26.
- WALKER, E. L. Action decrement and its relation to learning. *Psychological Review*, 1958, **65**, 129-142.
- WALKER, E. L., & TARTE, R. D. Memory storage as a function of arousal and time with homogeneous and heterogeneous lists. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 113-119.

NOTE

1. An excellent example of the problems (experimental and theoretical) of exact and partial replications is a set of experiments which sought to determine an influence of warm-up on recall (Dinner & Duncan, 1959; Irion, 1949; Lazar, 1967; Rockway & Duncan, 1952). There, even the failure of an exact replication (1952) did not settle concern over the topic.

(Received for publication September 15, 1976.)