Psychonomic Bulletin & Review
2008, 15 (2), 302-308
doi: 10.3758/PBR.15.2.302

Semantic organization of study materials has opposite effects on recognition and recall

SCOTT A. GUERIN AND MICHAEL B. MILLER
University of California, Santa Barbara, California

It has been well established for several decades that semantic organization of study materials greatly enhances recall by facilitating access to information during retrieval. However, the effect of organization on recognition, and its relationship to the effect on recall, is in doubt. We report the first direct comparison of the effects of categorically organizing study lists on recognition, cued recall, and free recall. We found that whereas organization improved recall, it impaired recognition. Organization had a larger effect on free recall than on cued recall. Within the categorized lists, recall was superior for items highly associated with the category; the opposite was true of recognition. In recall, organization improved the proportion of categories recalled, but it lowered the proportion of items per category recalled. A simple framework for interpreting the dissociation is offered. Possible mechanisms underlying the detrimental effect of organization on memory and prospects for future research are briefly discussed.

It is well documented that semantic organization of study materials greatly enhances recall. Participants recall a list grouped into categories better than the same list randomly ordered, and recall of an item in a category increases according to its association with the category (Cofer, Bruce, & Reicher, 1966). In free recall, participants tend to cluster their responses categorically (Cofer et al., 1966; Cohen, 1963). In cued recall, presenting a category label greatly enhances memory for a categorized list (Tulving & Pearlstone, 1966). It seems likely that organization facilitates recall because it facilitates access to information during retrieval by providing a potent, easily accessible cue and an easily implemented, effective retrieval strategy (Mandler, 1972; Tulving & Pearlstone, 1966).

Although it is well established that semantic organization aids recall, the effect of organization on recognition has been much more difficult to ascertain. There are reports that organization has no effect on recognition (Bruce & Fagan, 1970; Kintsch, 1968), that organization facilitates recognition (Connor, 1977; D'Agostino, 1969; Mandler, 1972; Neely & Balota, 1981), and that organization impairs recognition (Hintzman, 1988; Shiffrin, Huber, & Marinelli, 1995). This divergence of results is undoubtedly due to the substantial variations in procedural methods across studies. Nevertheless, much of the recent work on this problem constitutes a methodologically cohesive body of work that clearly suggests that semantic organization can impair recognition. This work points to a need to revisit the classic problem of the differential effect of organization on recognition and recall originally posed by Kintsch (1968). However, the discrepancies in the previous work must be borne in mind throughout. Our hypothesis is not that organization impairs recognition under any and all circumstances, but rather that organization can impair recognition under a broad range of circumstances, and that it can impair recognition under the same conditions in which it aids recall.

Many of the more recent studies on recognition have employed a similar methodology. These studies manipulated organization by varying the number of items per category, and employed an orientation task to constrain the participants' processing of stimuli during encoding. For instance, Hintzman (1988) varied the number of items per category in a randomly ordered list of constant length. He observed that, as the number of items per category increased, both the false alarm rate and the hit rate increased. However, the increase in the false alarm rate greatly outpaced the increase in the hit rate, resulting in an overall decline in recognition performance. He successfully accounted for his results with a global matching model (see also Shiffrin et al., 1995). Similar results have been obtained by Shiffrin et al. (1995) and Dewhurst (2001) using word lists, and by Koutstaal and Schacter (1997) using pictures (see also Criss & Shiffrin, 2004).

These studies suggest that semantic organization of study materials can actually impair recognition. This could result from a reliance on "gist-based" memory, leading to a disproportionate increase in the false alarm rate (Brainerd & Reyna, 2002). It could also result from a decrease in the novelty or distinctiveness of items, which can impair memory (Hunt, 1995; Tulving & Kroll, 1995). In recall, these detrimental processes may be overshadowed by the facilitation in the endogenous production of retrieval cues that semantic organization provides. But in recognition,

S. A. Guerin, guerin@psych.ucsb.edu

the burden on the participant to endogenously produce retrieval cues is greatly diminished by the availability of a potent cue in the external world: the to-be-remembered word itself. If this line of reasoning is correct, then semantic organization of study materials may constitute a new dissociation between recognition and recall.

To test this hypothesis, we compared the effects of organization on three memory tests: recognition, cued recall, and free recall. We assume that the burden on the participant to endogenously produce retrieval cues is intermediate in the cued recall condition. We evaluated the effect of organization in two ways: (1) by comparing categorized and uncategorized lists and (2) by studying the effect of the degree to which an exemplar was associated with a categorized list. We anticipated that organization would facilitate free recall and hinder recognition. Critically, we also hypothesized that the effect on cued recall would be intermediate between the effects on recognition and free recall.

METHOD

Participants

A total of 108 students at the University of California, Santa Barbara were given course credit for their participation. Each was randomly assigned to recognition, cued recall, or free recall. Data from 2 participants were replaced because of poor performance. All participants gave informed consent, as approved by the Institutional Review Board at the University of California, Santa Barbara.

Stimuli

We selected 36 categories from the Van Overschelde, Rawson, and Dunlosky (2004) norms. The first 20 exemplars were taken from each category. Occasionally, however, an exemplar was excluded if it could also be considered a member of another category being used. We created a 10-point ordinal scale, referred to as *category rank*, that represented the frequency with which the word was produced as a member of the category in the normative data. Low values of category rank correspond to items that are closely associated with the category. The first 2 exemplars were assigned a category rank of 1, the next 2 exemplars were assigned a category rank of 2, and so forth. For each participant, 6 categories were used to form six categorized lists of 10 items each, with each level of category rank represented once. Two exemplars from each of the 30 remaining categories were randomly selected without regard to rank and assigned to six uncategorized lists of 10 items, with the constraint that 2 items from the same category did not appear in the same list. Stimuli were rotated through conditions so that each word occurred equally often in each condition.

Because we tested memory for all studied items in the recognition test, so as to maximize the number of observations, the categorized and uncategorized conditions differed in terms of number of items per category presented during both encoding and retrieval. Studies that have held the number of items per category presented in a recognition test constant have obtained results similar to ours (Dewhurst, 2001; Hintzman, 1988; Koutstaal & Schacter, 1997; Shiffrin et al., 1995).

Procedure

Participants were tested individually and were aware at the outset that they were participating in a memory experiment. During the study phase, the lists were presented in a random order and without interruption. Items within each list were also presented in a random order. Thus, items from each list were presented in a blocked format. Each word appeared on a computer screen for 2 sec. The participants indicated by a keypress whether they liked or did not like each word. Following the study session, the participants counted backward by 3 for 30 sec. Then the participants were given the memory test. For the recognition group, each study item plus an equivalent number of same-category distractors were presented in a random order. Participants indicated by a keypress whether each word was *old* or *new*. Following the recognition decision, a 5-point confidence scale appeared, with 1 labeled as *guessing* and 5 labeled as *certain*. Participants indicated their confidence in the recognition judgment by a keypress. All decisions were self-paced. In the cued recall group, participants were presented with the names of all 36 categories in a random order on the computer screen. Participants were told that the words they studied fell into general categories, that these were the names of those categories, and that they should use these names to help them remember the items. They were given sheets on which to record their responses. The participants rated their confidence in each response after writing it down. Participants were given 8 min to complete the recall test. The free recall group was identical to the cued recall group except that they were not given the category names. The recall false alarm rate was determined by identifying all intrusions that occurred in the stimulus norms and assigning them to the condition in which that category was represented.

RESULTS

Categorized Versus Uncategorized Lists

The hit and false alarm rates for the categorized and uncategorized lists are reported for each of the three groups in Figure 1.

Recognition. As expected, categorization produced a substantial increase in the false alarm rate (.25 vs. .10). The effect on the hit rate was negligible (.89 vs. .87). Thus, there was an overall decline in memory performance in the categorized condition. An ANOVA indicated a significant interaction between categorization (categorized vs. uncategorized) and study (targets vs. distractors) [$F(1,35) = 67.73$, $p < .001$]. Paired t tests indicated that whereas categorization had a significant effect on the false alarm rate [$t(35) = 10.06$, $p < .001$], it had no detectable effect on the hit rate [$t(35) = 1.59$, $p = .12$].

Because the hit rates were rather high, there was some concern that a ceiling effect had occurred. However, when we restricted attention to "old" responses receiving the highest confidence rating, the mean hit rate fell to .78 (see Table 1). Nevertheless, a parallel analysis conducted on these data produced the same conclusions as those reported throughout.

The mean of each participant's median recognition reaction times (RTs) are reported in Figure 2. RTs were longer for distractors than for targets [$F(1,35) = 85.24$, $p <$ .001]. Of greater interest, RTs were longer for categorized items than for uncategorized items [$F(1,35) = 17.48$, $p <$ .001]. The interaction was significant [$F(1,35) = 14.81$,

Table 1
Recognition Hits and False Alarms at the Highest Level of Confidence

	Hits		False Alarms	
	M	*SEM*	*M*	*SEM*
Categorized	.77	.009	.11	.007
Uncategorized	.80	.006	.04	.004

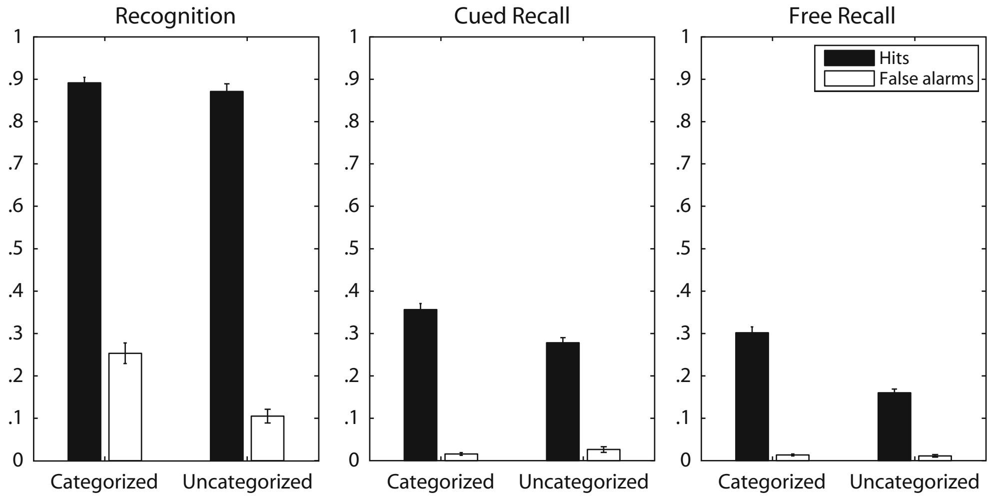

Figure 1. Mean hit and false alarm rates (± *SEM*) as a function of categorization and memory test.

$p < .001$]. Whereas the effect of categorization on distractors was significant [$t(35) = 5.69$, $p < .001$], the effect on targets was not [$t(35) = 1.51$, $p = .14$]. The effect was marginally significant when only correct responses were considered [$t(35) = 2.08$, $p = .05$].

Normalized receiver operating characteristics (z-ROCs) based on the group mean hit and false alarm rates are shown in Figure 3. The categorized and uncategorized z-ROCs fall on two distinct lines. On the basis of the fits in Figure 3, the sensitivity measure d_a was 1.98 for the categorized condition and 2.18 for the uncategorized condition (Macmillan & Creelman, 1991). We also estimated z-ROCs individually for each participant (mean $r^2 \pm SEM$ was $.82 \pm .03$ for the categorized condition and $.78 \pm .04$ for the uncategorized condition; we corrected for values of 0 and 1 by replacing them with $1/(2N)$ or $1 - 1/(2N)$, as necessary; see Macmillan & Creelman, 1991). The mean $d_a \pm SEM$ was 2.10 ± 0.12 for the categorized condition and 2.34 ± 0.13 for the uncategorized condition. This difference was significant [$t(35) = 2.21$, $p = .03$]. A similar result was obtained when individual z-ROCs were computed using the slope from the group plot. These analyses are consistent with the conclusion that categorization impairs recognition performance.

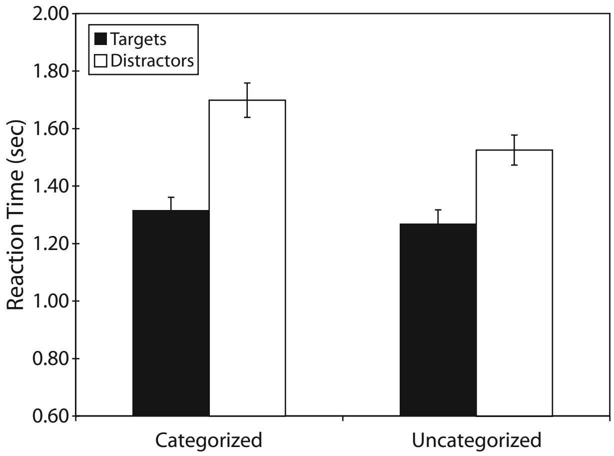

Figure 2. Group mean of each participant's median reaction time (±*SEM*) on the recognition test as a function of categorization and whether the test item was a target or a distractor.

Estimates of criteria also differed between conditions. On the basis of the fits in Figure 3, c_2 was $-.21$ for the categorized condition and .04 for the uncategorized condition. Estimated individually for each participant, this difference was significant [categorized, $-.23 \pm .04$; uncategorized, $.04 \pm .02$; $t(35) = 7.00$, $p < .001$].

Recall. Both forms of recall showed a substantial increase in the hit rate in the categorized condition. Intrusions were rare across the board, as is generally observed. Because the false alarm rates were at floor, the variance was very low, and it was thus inappropriate to include the false alarm rate alongside the hit rate in an ANOVA. Paired t tests indicated that, for both forms of recall, the increase in the hit rate was significant [cued recall, .36 vs. .28, $t(35) = 4.30$, $p < .001$; free recall, .30 vs. .16, $t(35) = 8.73$, $p < .001$]. The effect on the hit rate was larger in free recall than in cued recall [$F(1,70) = 6.23$, $p = .02$]. There was no effect of categorization on the false alarm rate [cued recall, categorized = .02, uncategorized = .03, $t(35) = -1.54$, $p = .13$; free recall, .01 vs. .01, $t(35) = .45$, $p = .66$]. This is an obvious floor effect, but intrusions are usually rare in recall; thus, this effect likely reflects an enduring property of recall rather than an idiosyncrasy of this experiment.

Intrusions that were not in the stimulus norms were not included in the recall false alarm rate. In cued recall, there were 35 such intrusions (out of a total of 1,494 responses). Of these, 15 appeared to be related to categories assigned to the categorized condition and 16 appeared to be related to categories assigned to the uncategorized condition, on the basis of the determination of a research assistant. In free recall, there were 36 of these intrusions (out of a total of 1,094 responses). All of these appeared to be unrelated to the studied categories.

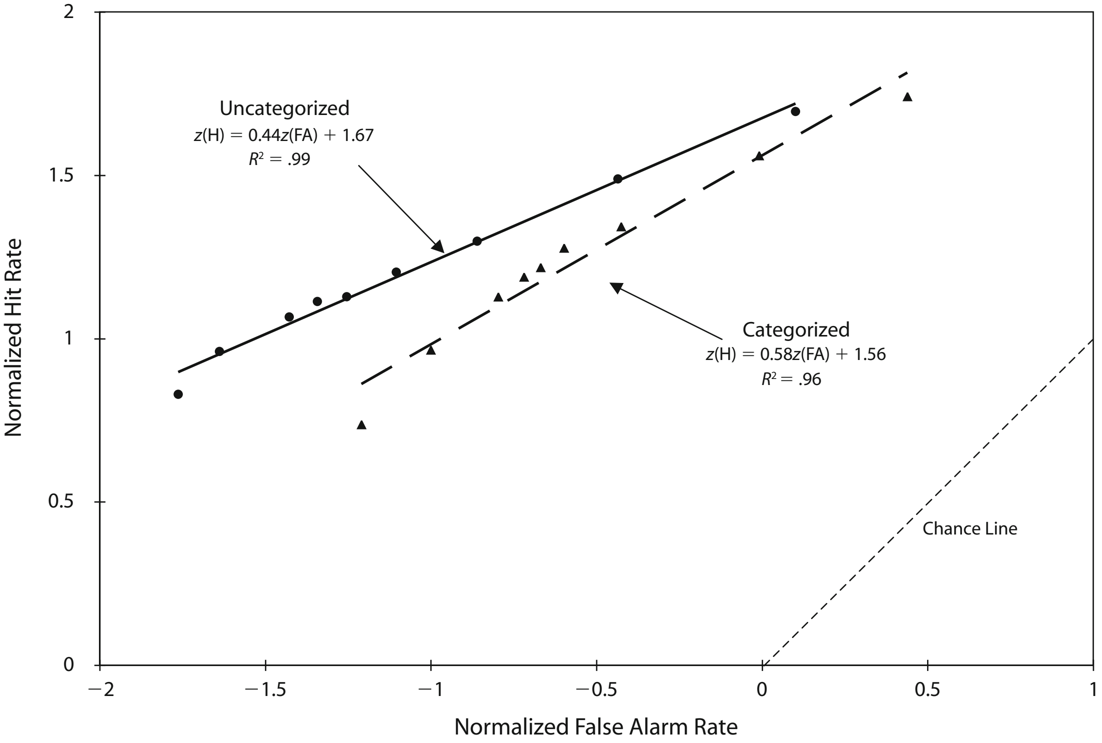

Figure 3. Normalized receiver operating characteristics plotted in *z*-space for the categorized and uncategorized conditions based on the group-averaged confidence rating data from the recognition test. Lines show least squares regression.

Category Rank

The hit and false alarm rates are shown as a function of category rank in Figure 4. Low values of category rank correspond to items that are closely associated with a category.

Recognition. In recognition, category rank had a large effect on the false alarm rate but a negligible effect on the hit rate. The false alarm rate was approximately halved across the entire range of category rank. An ANOVA revealed a significant interaction between study and rank [$F(9,315) = 3.96$, $p < .001$]. To investigate the interaction, ANOVAs were computed separately for targets and distractors. There was no effect of rank on targets [$F(5.82, 203.80) = 1.80$, $p = .10$]. Throughout, noninteger degrees of freedom indicate a Greenhouse–Geisser correction. However, there was an effect of rank on distractors [$F(9,315) = 6.75$, $p < .001$].

Recall. For both forms of recall, there was a strong, negative relationship between category rank and the hit rate [cued recall, $F(9,315) = 10.89$, $p < .001$; free recall, $F(9,315) = 9.56$, $p < .001$]. For both forms of recall, there was also a negative relationship between the false alarm rate and category rank, but the magnitude of the effect was smaller [cued recall, $F(3.41,119.36) = 2.62$, $p = .047$; free recall, $F(3.63,127.12) = 3.13$, $p = .02$].

Number of Categories Recalled and Number of Words per Category Recalled

Following Cohen (1963), we evaluated two measures of recall: number of categories recalled and number of items per category recalled. The proportion of words recalled is equal to the proportion of categories recalled multiplied by the proportion of items within categories recalled. We define a category as "recalled" if at least one item from that category is correctly recalled. The results of this tabulation are shown in Table 2. ANOVAs were computed separately for the two measures. There are two critical results. First, free recall and cued recall differed only in terms of proportion of categories recalled [$F(1,70) = 74.66$, $p < .001$], not proportion of items per category recalled [$F(1,70) = 0.17$, $p = .68$]. Second, categoriza-

Table 2
Proportion of Categories Recalled and Proportion of Words per Category Recalled

	Categories Recalled				Words per Category Recalled			
	Categorized		Uncategorized		Categorized		Uncategorized	
Group	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
Cued recall	.90	.02	.47	.02	.40	.01	.58	.01
Free recall	.74	.02	.28	.01	.41	.02	.59	.02

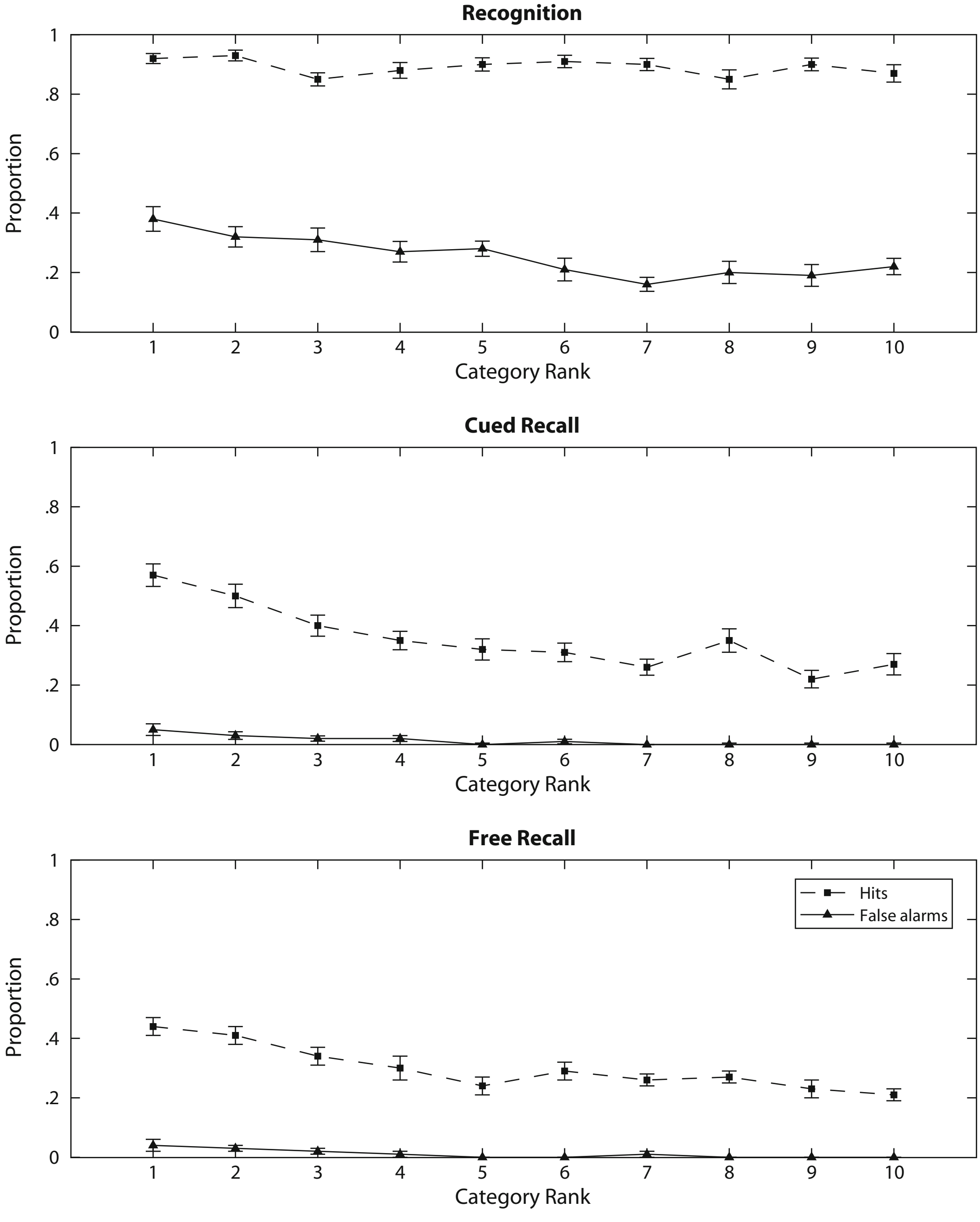

Figure 4. Mean hit and false alarm rates (± *SEM*) as a function of category rank and memory test. Low values of category rank correspond to items that are closely associated with a category and high values of category rank correspond to items that are more distantly associated with a category.

tion had two effects: (1) It increased the proportion of categories recalled [$F(1,70) = 512.17$, $p < .001$], and (2) it decreased the proportion of items per category recalled [$F(1,70) = 151.32$, $p < .001$].

DISCUSSION

We investigated the effect of semantic organization of study materials on recognition, cued recall, and free recall. We found that whereas organization impaired recognition, it facilitated recall. To our knowledge, this is the first study to directly compare the effects of organization on recognition, cued recall, and free recall. It is also, to our knowledge, the first study to show that, under the same encoding conditions, organization can hinder recognition but aid recall.

Several lines of evidence suggest that organization impaired recognition. Whereas the false alarm rate increased substantially in the categorized condition, the hit rate was unaffected. Within the categorized condition, category rank had a large effect on the false alarm rate, but no detectable effect on the hit rate. Measures of sensitivity

(d_a) were greater for the uncategorized condition than for the categorized condition. Taken together, these findings strongly suggest that semantic organization of study materials impaired recognition memory in this experiment.

In recall, categorization caused a large increase in the hit rate. There was not a significant effect on the false alarm rate. Therefore, we replicated the classic result that recall is superior for organized study lists. Within the categorized condition, items closely associated with the category were much more likely to be correctly recalled. However, there was also a small but significant tendency for items closely associated with the category to be falsely recalled. This result points to a similarity between recognition and recall, and suggests that the same tendency for semantic organization to elicit errors of commission is still present in recall but is greatly attenuated.

Although we clearly observed a decline in recognition as a result of organization, one can imagine conditions under which organization could facilitate recognition. For instance, such a result could have been obtained if participants actively generated their own organizational scheme (Mandler, 1972), if participants were led to anticipate a recall test and engaged in unconstrained encoding (Neely & Balota, 1981), or if the ordering of test items closely approximated the ordering of study items, thus encouraging interitem contextual cuing during recognition (Jacoby, 1972). Our conclusion is not that organization impairs recognition under any and all circumstances, but rather that organization can impair recognition under a broad range of circumstances, and that it can impair recognition under the same conditions in which it aids recall.

It is important to bear in mind that comparisons between recognition and recall raise a number of interpretive difficulties. In particular, performance in recognition may not be linearly related to performance in recall. As a result, direct quantitative comparisons and interpretations of certain interactions are difficult (Brown, 1976; Loftus, 1978). Here, we are merely making claims about the relative direction of the effect of organization.

Why should organization impair recognition under the same conditions in which it aids recall? An organized list is one in which items have been grouped according to their similarity along some dimension. Thus, items within organized lists are more similar and thus less novel and distinctive. It is known that when an item in a list is unique, it is better remembered (Hunt, 1995). It is also known that items that are novel are better remembered than items that are familiar (Tulving & Kroll, 1995). For these reasons, we should expect that, in general, items within an organized list should not obtain the same level of item-specific memory. Accordingly, we observe that recognition performance is poorer for categorized lists and that the number of items per category recalled is less for categorized lists than for uncategorized lists.

It is also known that participants tend to abstract general characteristics of objects and events that are encountered, often referred to as "gist" (Brainerd & Reyna, 2002). Even when specific details of an event are lost, general characteristics of the event may be retained, and these generalizations about the event may be used to infer lost details. The tendency to rely on gist is likely to dominate when participants base their decisions on a vague familiarity with items, rather than retrieval of contextual details (Yonelinas, 2002). Furthermore, one expects the false alarm rate to be more sensitive to these familiarity effects than the hit rate (see also Koutstaal & Schacter, 1997). As a result, we observe a very large increase in the recognition false alarm rate for categorized lists. Thus, organization of study materials hinders memory in two ways: first, by decreasing the distinctiveness and novelty of individual items; second, by producing familiarity-induced errors of commission.

But organization can also facilitate memory. Organization can provide a powerful, easily accessed cue during recall. Even in the absence of a specific external cue, participants recalled most of the categories from the categorized condition (.74). When participants in the free recall condition did remember at least one item from a category, they recalled the same average number of items from that category (.41) as did the cued recall group (.40), which replicates previous findings (Tulving & Pearlstone, 1966). The items they recalled tended to be those that were highly associated with the category. This implies that the function of organization in recall is to cue memory for items related to organizational structures, and that the major challenge in free recall is to access the organizational cues themselves.

Our results shed light on the longstanding problem of the differential effect of semantic organization on recognition and recall. Semantic organization can hinder recognition under the same conditions in which it facilitates recall. This observation adds to a short list of experimental manipulations that dissociate recognition and recall. Further exploration of the impairment in recognition, in terms of both distinctiveness effects and the differential roles of familiarity and recollection, is a compelling avenue for future work. Our observations also bring new insights to bear on the problem of false recognition by demonstrating that the same manipulations that produce robust false recognition can also provide for a substantial benefit in recall. Indeed, the heavy reliance of recall on the availability of organizational information may be a major reason why human memory places such a premium on the extraction of gist even though doing so often comes at a severe cost.

AUTHOR NOTE

We thank Denise Wernli, Chris Jensen, and Rachel Segal for assistance and Stan Klein for helpful comments. Correspondence concerning this article should be addressed to S. A. Guerin, Department of Psychology, University of California, Santa Barbara, CA 93106-9660 (e-mail: guerin@psych.ucsb.edu).

REFERENCES

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, **11**, 164-169.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and recognition* (pp. 1-35). London: Wiley.

Bruce, D., & Fagan, R. L. (1970). More on the recognition and free recall of organized lists. *Journal of Experimental Psychology*, **85**, 153-154.

Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in

free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, **71**, 858-866.
Cohen, B. H. (1963). An investigation of recoding in free recall. *Journal of Experimental Psychology*, **65**, 368-376.
Connor, J. M. (1977). Effects of organization and expectancy on recall and recognition. *Memory & Cognition*, **5**, 315-318.
Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, **111**, 800-807.
D'Agostino, P. R. (1969). The blocked-random effect in recall and recognition. *Journal of Verbal Learning & Verbal Behavior*, **8**, 815-820.
Dewhurst, S. A. (2001). Category repetition and false recognition: Effects of instance frequency and category size. *Journal of Memory & Language*, **44**, 153-167.
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.
Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, **2**, 105-112.
Jacoby, L. L. (1972). Effects of organization on recognition memory. *Journal of Experimental Psychology*, **92**, 325-331.
Kintsch, W. (1968). Recognition and free recall of organized lists. *Journal of Experimental Psychology*, **78**, 481-487.
Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory & Language*, **37**, 555-583.
Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, **6**, 312-319.
Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
Mandler, G. (1972). Organization and recognition. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 139-166). New York: Academic Press.
Neely, J. H., & Balota, D. A. (1981). Test-expectancy and semantic-organization effects in recall and recognition. *Memory & Cognition*, **9**, 283-300.
Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 267-287.
Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, **2**, 387-390.
Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, **5**, 381-391.
Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory & Language*, **50**, 289-335.
Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, **46**, 441-517.

(Manuscript received November 21, 2006;
revision accepted for publication August 23, 2007.)