Novelty assessment in the brain and long-term memory encoding

ENDEL TULVING
University of California, Davis, California
and Rotman Research Institute, Toronto, Ontario, Canada

and

NEAL KROLL
University of California, Davis, California

Recent positron emission tomography (PET) studies have identified neuronal components of widespread novelty-assessment networks in the brain. We propose that the efficacy of encoding on-line information into long-term memory depends on the novelty of the information as determined by these networks, and report a test of this "novelty/encoding" hypothesis. Subjects studied a list of words. Half of the words were "familiar" by virtue of their repeated presentation to the subjects before the study of the critical list; the other half were novel, in that they had not previously been encountered in the experiment. The results conformed to the prediction of the novelty/encoding hypothesis: accuracy of explicit (episodic) recognition was higher for novel than for familiar words.

Previous research using the technique of positron emission tomography (PET) has indicated that the human brain contains widely distributed novelty-detection networks. Cortical and subcortical regions in the expanded limbic system as well as in the temporal lobes respond more actively to novel stimuli than to otherwise comparable familiar stimuli (Tulving, Markowitsch, Kapur, Habib, & Houle, 1994). These findings with humans are consistent with the results of single-unit recording studies that have revealed the existence of neurons in the homologous regions of monkey brains that respond more vigorously to novel than to familiar or recently seen stimuli (Fahy, Riches, & Brown, 1993; Li, Miller, & Desimone, 1993; Riches, Wilson, & Brown, 1991; Wilson & Rolls, 1993).

The discovery of specific neuronal networks in the brain that are differentially involved in the processing of novel rather than familiar perceptual inputs makes excellent sense in light of massive behavioral evidence showing that all animals routinely respond differentially to novel versus familiar stimuli. These networks also make good evolutionary sense. Distinguishing between what is novel and what is already known, and distinguishing between degrees and kinds of novelty, is one of the elemental requirements for reacting adaptively to the happenings in one's proximate environment.

One function served by the neuronal novelty-assessment networks is to determine the necessity of encoding the information for long-term storage. We conjecture that the encoding of incoming information for long-term storage depends on the novelty of the on-line information. A stronger form of this conjecture is that novelty is a necessary, although not a sufficient, condition for the long-term storage of information.

We will refer to this conjecture as the "novelty/encoding hypothesis." This hypothesis is rooted in an earlier suggestion that left-frontal cortical regions are involved in the encoding of novel information for episodic memory storage (Tulving, Kapur, Craik, Moscovitch, & Houle, 1994). Here we add the suggestion that novelty of information is determined by neuronal networks whose output provides the necessary information for the frontal encoding networks.

Thus, the novelty/encoding hypothesis holds that the novelty-assessment mechanism is a component of encoding processes. The neuronal novelty-assessment circuits serve the function of identifying on-line information whose encoding for long-term storage is of adaptive significance, and transmit it for further processing. Similar ideas have been proposed by Fabiani and Donchin (1995), Kohonen, Oja, and Lehtiö (1989), Metcalfe (1993), Siddle, Packer, Donchin, and Fabiani (1991), and Sokolov (1963), among others.

According to the novelty/encoding hypothesis, then, encoding processes, as traditionally conceptualized (Craik & Lockhart, 1972; Tulving, 1983), consist of two sets of concatenated subprocesses: (1) novelty assessment, subserved by subcortical and cortical neuronal networks in the limbic system and temporal lobes, and (2) higher level (contextually determined) encoding operations, subserved by cortical regions that include left frontal lobes. The end product of the concatenation is the engram or memory trace.

On-line information, including that provided by the sensory systems, is transmitted for higher level encoding on the basis of its novelty: novel information receives preferential treatment over familiar information. Completely redundant information is screened out from further processing and is not stored.

We report a simple experiment that was designed as a direct test of the novelty/encoding hypothesis. Normal adult subjects were given an explicit recognition test for words varying only with respect to their novelty and familiarity. All other variables were held constant. The prediction made from the novelty/encoding hypothesis was straightforward: Explicit recognition of studied words—that is, identification of words as having been presented in a particular study episode—is inversely related to the words' prestudy familiarity.

METHOD

Design

The design was a 2×2 within-subjects factorial, as shown in Table 1: two levels of prestudy familiarity (novel or familiar) were combined with two kinds of test words in the yes/no recognition test (old and new). Four different subsets of materials were rotated through the four experimental conditions. Thus, 4 subjects provided for one full instantiation of the experiment free from material-specific confoundings. We tested 4 university-educated subjects, 2 women and 2 men.

Materials

A total of 160 English nouns of two syllables constituted the basic pool of experimental words. This pool was divided into four subsets: A, B, C, and D, each of 40 words. For any given subject, two of these sets (e.g., A and B) were used in prefamiliarized trials of the presentation of two subsets of words.

Procedure

Familiarization. Two sets of items (80 total) were first presented four times. The 320 tokens appeared on the screen of a computer monitor, one at a time and at a subject-determined rate, which was typically approximately 1.5 sec per word. The subjects were asked to make living/nonliving decisions about the words. At the end of the presentation, the subjects were given a yes/no recognition test of the 80 words without any distractors. Subjects' recognition performance revealed very few "no" responses. This test was followed by one more trial of living/nonliving judgments on the 80 "familiarized" words. The order of words was individually randomized for each subject and each trial, using a computerized randomization program. The subjects saw familiarized words a total of six times.

Interpolated task. Following the familiarization phase, the subjects were asked to spend 5 min typing the names of as many cities of the world as they could on the computer keyboard.

Critical study. The subjects were shown 80 words. Half had appeared in the familiarization phase, and half were novel. The sub-

Table 1
Design of the Experiment

Words	Studied	Nonstudied
Novel	H_N	FA_N
Familiar	H_F	FA_F

Note—H and FA represent hit rates and false-alarm rates, respectively; subscripts N and F indicate the experimentally defined novelty/familiarity status of the test words. The prediction based on the novelty/encoding hypothesis is that $H_N - FA_N > H_F - FA_F$.

jects were told that this was the "critical" list and that their memory for the words in this list would be tested. They were also told that they had already seen some of the critical study words before, but had not encountered others in the experiment. During the presentation of the words for study, they were again asked to make the living/nonliving judgments. Finally, they were told that they should do their best to impress each and every one of the shown words on their minds, so that they could later remember the words in this "critical study" list. As before, the subjects controlled the rate of presentation, which on average was about 2–3 sec per word.

Interpolated task. The study of the critical list was followed by the subjects' typing of the names of celebrities and famous people for 5 min.

Test. The subjects were shown all 160 words from the pool, one at a time at a self-paced rate, and asked to indicate for each word (1) whether or not it had been on the critical study list ("yes"/"no") and (2) their confidence in these judgments: 3—*very confident*, 2—*more or less sure*, and 1—*guessing*.

Logic of the Study and Treatment of the Data

We manipulated the preencoding novelty/familiarity of common words. The novel words had been seen and heard by the subjects outside the laboratory a large number of times, but they had not been seen or heard in the laboratory before their appearance in the critical study list. The familiar words had also been encountered outside the laboratory a large number of times but, in addition, had also been encountered in the laboratory before their appearance in the critical study list. A similar study was reported by Kinsbourne and George (1974).

The data of interest consist of the number (or proportion) of items identified by the subjects as the critical study-list words in each of the four conditions of the experiment (Table 1) and means of the confidence ratings.

RESULTS AND DISCUSSION

The results are summarized in Table 2. They confirmed the prediction of the novelty/encoding hypothesis: Recognition accuracy was greater for novel preencoding words than for familiar preencoding words. When recognition of study-list words was assessed in terms of the differences between hit and false-alarm rates for the critical study list, recognition accuracy was .20 for familiar words and .56 for novel words. Analysis of variance of these proportions yielded a significant interaction between the novelty/familiarity status of test words and their appearance or nonappearance in the critical study list [$F(1,3) = 280.3, MS_e = .069, p < .01$]. The data also showed that subjects were more confident in their correct identifications of novel-word appearance or nonappearance in the critical list (mean difference of 2.81 on the 6-point scale) than they were in their judgments of familiar words [mean difference of 0.68; the interaction $F(1,3) = 42.6, MS_e = .317, p < .01$].

The conclusion that discrimination of test items as having or not having occurred in the critical study list was higher for novel than for familiar words follows directly from the experimental design and the observed results. Because all other variables, including the higher level (semantic) encoding operations at study, were held constant in the experiment, the initial source of the observed difference in discrimination must lie in the novelty/familiarity status of the critical study-list words. The outcome is robust. Using lists only half as long as those used here, only

Table 2
Proportion of Test Words Judged to be Critical Study-List
Words With Mean Confidence Ratings (in Parentheses)

	Studied	Nonstudied
Novel	.77 (1.54)	.21 (−1.27)
Familiar	.67 (0.70)	.47 (0.02)

three continuous familiarization trials, and only a 2-min interpolated activity between the critical study list and the recognition test, we have done a simpler version of the experiment, and obtained a pattern of data very similar to that reported here.

The findings of our experiment could have been readily predicted on the basis of available evidence. A long string of experiments have shown that words less frequently encountered in language are better recognized in episodic-memory experiments than are more frequently occurring words (e.g., Balota & Neely, 1980; Glanzer & Bowles, 1976; Gregg, 1976). It is also known that experimental pre-exposure of targets and lures reduces the recognition of subsequently studied words (Kinsbourne & George, 1974), and that repeated presentation of subsets of words from a larger experimental set on successive learning trials reduces subjects' ability to explicitly recognize words from the most recent trial (Anderson & Bower, 1972). In light of available evidence, the data we present are not exactly newsworthy.

What may be somewhat more novel is the suggestion that our findings reflect the operations of the novelty assessment mechanisms of the brain: Explanations of psychological findings in terms of brain function have not been terribly popular in cognitive psychology. We propose that familiar items are less well recognized than novel items because the novelty-assessment system screens out familiar items from further processing for subsequent recognition at an early stage of encoding. The same screening mechanism may also provide a simple explanation, either partly or in full, for several other phenomena of recognition memory, including the following: (1) Novelty effects studied with event-related potentials (Knight, 1984, in press; Nelson & Collins, 1991); (2) habituation–novelty effects and novelty preferences by infants (Fagan, 1990; MacKay-Soroka, Trehub, Bull, & Corter, 1982; Rose & Slater, 1983); (3) correlation between distinctiveness and memorability of faces (Sommer, Heinz, Leuthold, Matt, & Schweinberger, 1995); (4) distinctiveness effects in explicit recognition tests (Jacoby, Craik, & Begg, 1979); (5) the isolation effect (Hunt, 1995); (6) flashbulb memories (Conway et al., 1994; Winograd & Neisser, 1993); (7) semantic satiation (Esposito & Pelton, 1971; Smith & Klein, 1990); (8) the generation effect (Slamecka & Graf, 1978); and (9) mnemonic benefits of perceptual interference (Hirshman, Trembath, & Mulligan, 1994).

Caveats

We have defined "novelty" and "familiarity" purely operationally, in terms of frequency and recency of prior occurrence. Neither term refers to a unitary entity or a single process. Novelty and familiarity assume a large number of different forms. Both are only general labels for sets of complex processes whose components remain to be identified. Schmidt's (1991) four classes of distinctiveness may provide a good starting point for such an undertaking. The question of novelty, or familiarity, of "exactly what?" must also be considered. Every "frame" in the stream of consciousness contains both novel and familiar features and aspects. The most common situation is the one in which objects familiar from the past appear in contexts that are novel in the present.

Although the present experiment was a typical episodic memory one, we have referred generally to "long-term" memory and not tried to distinguish between episodic and semantic memory. The basic idea of novelty/encoding may well apply to the uptake and storage of both episodic- and semantic-memory information. It is too early to tell.

We believe that novelty/encoding applies to long-term memory only. The operations of the novelty-assessment system probably do not affect attentional or working-memory processing of on-line information. It is perfectly possible to pay attention to, and to hold in primary memory, any incoming information regardless of its familiarity.

Finally, the pattern of data we have reported from our experiment holds for recognition, whether free choice or forced choice. It need not hold for free recall. We have some preliminary data showing that familiar items may be more readily recallable than novel items. We are currently exploring the matter as an extension of earlier related work by others (Anderson & Bower, 1972; Mandler, Goodman, & Wilkes-Gibbs, 1982).

Conclusion

When all other conditions are held constant, experimentally novel words from a studied list show higher accuracy of explicit recognition than do familiar words. We propose that these, as well as related data reported by others, reflect the natural consequences of the operation of novelty-assessment networks in the brain.

REFERENCES

ANDERSON, J. R., & BOWER, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79, 97-123.

BALOTA, D. A., & NEELY, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 576-587.

CONWAY, M. A., ANDERSON, S. J., LARSEN, S. F., DONNELLY, C. M., McDANIEL, M. A., McCLELLAND, A. G., RAWLES, R. E., & LOGIE, R. H. (1994). The formation of flashbulb memories. *Memory & Cognition*, 22, 326-343.

CRAIK, F. I. M., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, 11, 671-684.

ESPOSITO, N. J., & PELTON, L. H. (1971). Review of the measurement of semantic satiation. *Psychological Bulletin*, 75, 330-346.

FABIANI, M., & DONCHIN, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 224-240.

FAGAN, J. F. (1990). The paired-comparison paradigm and infant intelligence. In A. Diamond (Ed.), *The development and neural bases of higher cognitive functions* (Annals of the New York Academy of

Sciences, Vol. 608, pp. 337-364). New York: New York Academy of Sciences.

FAHY, F. L., RICHES, I. P., & BROWN, M. W. (1993). Neuronal activity related to visual recognition memory: Long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. *Experimental Brain Research*, **96**, 457-472.

GLANZER, M., & BOWLES, N. (1976). Analysis of the word frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 21-31.

GREGG, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183-216). London: Wiley.

HIRSHMAN, E., TREMBATH, D., & MULLIGAN, N. (1994). Theoretical implications of the mnemonic benefits of perceptual interference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 608-620.

HUNT, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, **2**, 105-112.

JACOBY, L. L., CRAIK, F. I. M., & BEGG, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, **18**, 585-600.

KINSBOURNE, M., & GEORGE, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning & Verbal Behavior*, **13**, 63-69.

KNIGHT, R. T. (1984). Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalography & Clinical Neurophysiology*, **59**, 9-20.

KNIGHT, R. T. (in press). Distributed cortical network for visual stimulus detection. *Journal of Cognitive Neuroscience*.

KOHONEN, T., OJA, E., & LEHTIÖ, P. (1989). Storage and processing of information in distributed associative memory systems. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 129-167). Hillsdale, NJ: Erlbaum.

LI, L., MILLER, E. K., & DESIMONE, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, **69**, 1918-1929.

MACKAY-SOROKA, S., TREHUB, S. E., BULL, D. H., & CORTER, C. M. (1982). Effects of encoding and retrieval conditions on infants' recognition memory. *Child Development*, **53**, 815-818.

MANDLER, G., GOODMAN, G. O., & WILKES-GIBBS, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, **10**, 33-42.

METCALFE, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review*, **100**, 3-22.

NELSON, C. A., & COLLINS, P. F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events:

Implications for visual recognition memory. *Developmental Psychology*, **27**, 50-58.

RICHES, I. P., WILSON, F. A., & BROWN, M. W. (1991). The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. *Journal of Neuroscience*, **11**, 1763-1779.

ROSE, D. H., & SLATER, A. M. (1983). Infant recognition memory following brief stimulus exposure. *British Journal of Developmental Psychology*, **1**, 221-230.

SCHMIDT, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, **19**, 523-542.

SIDDLE, D. A. T., PACKER, J. S., DONCHIN, E., & FABIANI, M. (1991). Mnemonic information processing. In J. R. Jennings & M. G. H. Coles (Eds.), *Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches* (pp. 449-510). Chichester, U.K.: Wiley.

SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.

SMITH, L., & KLEIN, R. (1990). Evidence for semantic satiation: Repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 852-861.

SOKOLOV, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, **25**, 545-580.

SOMMER, W., HEINZ, A., LEUTHOLD, H., MATT, J., & SCHWEINBERGER, S. R. (1995). Metamemory, distinctiveness, and event-related potentials in recognition memory for faces. *Memory & Cognition*, **23**, 1-11.

TULVING, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press, Clarendon Press.

TULVING, E., KAPUR, S., CRAIK, F. I. M., MOSCOVITCH, M., & HOULE, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences*, **91**, 2016-2020.

TULVING, E., MARKOWITSCH, H. J., KAPUR, S., HABIB, R., & HOULE, S. (1994). Novelty encoding networks in the human brain: Positron emission tomography data. *NeuroReport*, **5**, 2525-2528.

WILSON, F. A., & ROLLS, E. T. (1993). The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. *Experimental Brain Research*, **93**, 367-382.

WINOGRAD, E., & NEISSER, U. (Eds.) (1993). *Affect and accuracy in recall: Studies of "flashbulb" memories*. New York: Cambridge University Press.