Research Report

The Generation Effect in Monkeys

Nate Kornell[1] **and Herbert S. Terrace**[2]

[1]*University of California, Los Angeles, and* [2]*Columbia University and New York State Psychiatric Institute*

ABSTRACT—***How well one retains new information depends on how actively it is processed during learning. Active attempts to retrieve information from memory result in more learning than passive observation of the same information (the* generation effect*). Here, we present evidence for the generation effect in monkeys. Subjects were trained to respond to five-item lists of photographs in a particular order. On some lists, they could request "hints" to guide their behavior; on others, they had to generate the correct order from memory. Training with hints resulted in high levels of initial performance, but accuracy dropped precipitously when the hints were removed on the criterion test. Training without hints led to relatively poor initial performance, but accuracy increased steadily and remained high on the criterion test.***

Psychologists have pondered the importance of how information is processed for more than a century. James (1890) remarked, "A curious peculiarity of our memory is that things are impressed better by *active* than by *passive* repetition" (p. 646, italics added). The *generation effect* refers to the fact that active retrieval of information from memory results in more learning than passive observation of the same information (Hirshman & Bjork, 1988; Jacoby, 1978; Slamecka & Graf, 1978). The generation effect is closely related to the testing effect—that is, the finding that taking tests during learning enhances long-term retention of information (Carrier & Pashler, 1992; Glover, 1989; Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a, 2006b).

The generation effect was discovered in experiments with human subjects who were trained in various verbal learning paradigms. In the present study, we asked whether an advantage of active learning is a fundamental principle of memory that can be studied in nonverbal animals. To address this question, we assessed the degree to which active processing affected serial learning in 2 rhesus macaque monkeys. Both subjects had extensive experience learning simultaneous chains, lists of arbitrarily selected items (usually photographs) displayed simultaneously on a touch-sensitive video monitor, in a different configuration on each trial (Terrace, Son, & Brannon, 2003). The subjects' task was to touch the items in a particular order, irrespective of their positions on the monitor. Each new list had to be learned by trial and error. Food reinforcement was provided only after the subject responded correctly to all of the list items. An error, at any position in the list, terminated the trial with a timeout.

Both subjects also had prior training in which a "hint" option was added to the simultaneous-chaining paradigm (Kornell, Son, & Terrace, 2007). Touching the hint icon caused a blinking border to appear around the list item to be selected next. Subjects learned to request hints. More important, the frequency with which they requested hints decreased as their accuracy increased. That relationship suggested that subjects were able to monitor their accuracy and request hints only when they were unsure about which item they should select next.

In the current experiment, each subject was trained on new five-item lists under four conditions. Under the first, the hint icon was not displayed, but hints were provided automatically on every trial. Under the second, subjects could obtain a hint on every trial by touching the hint icon. Under the third, they could obtain a hint on 50% of the trials; on the other 50%, the hint icon was not presented. Hints were never available under the fourth condition.

To assess the extent to which hints facilitated (or inhibited) learning during the first 3 days of training, we tested all subjects on the 4th day, without hints. If hints facilitated accuracy during training, and if a high level of accuracy during training indicates a high level of list learning, accuracy on this test would be expected to be higher on lists trained with hints than on lists trained without hints. Conversely, if the effort of generating correct responses is more effective than using a hint, accuracy on this test would be expected to be higher on lists trained without hints than on lists trained with hints.

Address correspondence to Nate Kornell, UCLA Department of Psychology, 1285 Franz Hall, Los Angeles, CA 90095-1563, e-mail: nkornell@psych.ucla.edu.

METHOD

Subjects
The subjects were two 10-year-old male rhesus macaques (Macduff and Oberon). They were maintained on a 12-hr light/dark cycle and housed individually in adjoining cages in a colony of 20 rhesus macaques at the New York Psychiatric Institute. The colony was maintained in accordance with National Institutes of Health guidelines. Subjects had ad lib access to water and were given daily rations of standard monkey chow (LabDiet®), fruit, and treats. Both subjects had extensive experience learning three-, four-, and seven-item lists as simultaneous chains (Terrace et al., 2003).

Apparatus
Subjects were tested 5 days per week in chambers that were housed in sound-attenuated booths. The chambers were made of stainless steel and tempered glass, and measured 23.0 in. high by 28.5 in. long by 27.0 in. wide. Each chamber was equipped with a Pixelink touch-sensitive video monitor (15-in. diagonal) that was controlled by a Macintosh G3 computer.

Procedure

Simultaneous-Chaining Paradigm
Five 1.5-in. × 2-in. photographs were displayed simultaneously on the monitor throughout each trial. The subject's task was to touch them in a particular, invariant order. The positions of the items varied randomly from trial to trial to ensure that subjects could not learn to execute the required sequence as a series of fixed motor responses. Correct responses were followed by brief (0.5 s) visual and auditory feedback. In addition, a correct response to the last item was followed by a food reward. An error ended the trial and initiated a 4-s time-out, during which the screen was dark.

The Hint Option
On hint trials, an icon was displayed on the right side of the monitor. Touching that icon caused four flashing lines to appear around the item to be selected next. A subject could request a hint for each list item. If a subject completed a trial correctly without any hints, he was rewarded with a highly desirable M&M candy. If the subject completed a trial correctly but requested one or more hints, he was rewarded with a less desirable 190-mg banana pellet. This contingency discouraged subjects from requesting a hint on every trial.

Design
Macduff and Oberon were trained and tested on 18 and 20 lists, respectively. Subjects were trained on each list for four consecutive sessions. Each session consisted of 60 trials of training on a single list.

Each list was assigned, during the first 3 days of training, to one of four conditions. In the *no-hint* condition, hints were never available. Accordingly, each list had to be learned entirely by trial and error. In the *100%-hint* condition, hints were available on every trial. In the *50%-hint* condition, a hint was available on only half of the trials. In the *auto-hint* condition, the hint icon was never available, but hints were provided automatically on every trial. How much a subject learned under each hint condition was assessed on the 4th day of training, when no hints were available.

Because the no-hint condition provided a baseline against which all the other conditions were measured, it accounted for approximately half of the lists on which the subjects were trained. Conditions were presented in an ABBA pattern with no-hint lists alternating with lists on which hints were presented. Macduff was trained on 8 no-hint, 5 50%-hint, 3 100%-hint, and 2 auto-hint lists; Oberon was trained on 10 no-hint, 5 50%-hint, 3 100%-hint, and 3 auto-hint lists.

RESULTS

Hint taking was defined as the relative frequency of trials on which a subject requested at least one hint; accuracy was defined as the relative frequency of trials on which subjects responded in the correct order to all five items. Assuming no backward errors (e.g., A-B-A), chance accuracy for completing a five-item list is .008 (.2 × .25 × .33 × .5).

Performance When Hints Were Available
Subjects requested hints on most of the trials on which they were available during the first 3 days of training. Oberon asked for hints on 98% and 95% of the 100%-hint and 50%-hint trials, respectively; Macduff asked for hints on 87% of both types of trials. Although hints were provided on 100% of the auto-hint trials, accuracy under that condition was not perfect. Macduff responded correctly on 69% of the trials; Oberon, on 96%.

At first glance, subjects' accuracy during the first 3 days of training suggests that they benefited from hints (see Fig. 1). However, these data cannot be used as evidence of learning because they include trials on which the subjects used hints. To obtain a measure of hints' effectiveness, we examined subjects' performance in the 50%-hint condition. When the hint was available, Oberon completed 80, 97, and 95% of the trials correctly on Days 1, 2, and 3, respectively; for Macduff, the corresponding percentages were 71, 79, and 69%. By contrast, subjects' accuracy on nonhint trials during 50%-hint training was almost nil. Oberon completed 1, 0, and 0% of the nonhint trials correctly on Days 1 through 3, respectively; Macduff completed 0, 3, and 9% of the nonhint trials correctly. Thus, when the hint was not available, subjects' accuracy barely exceeded the level predicted by chance, let alone levels one would

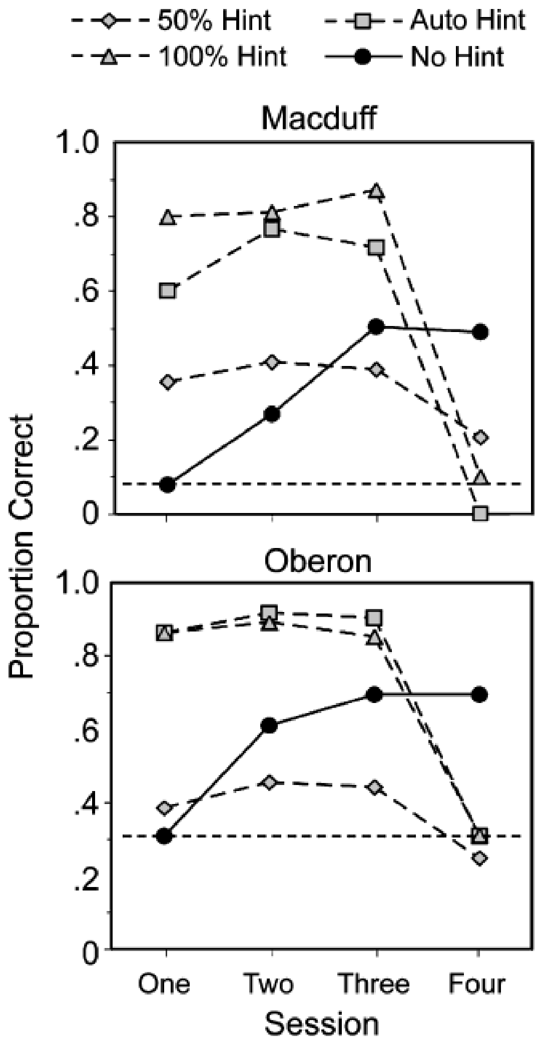

Fig. 1. Proportion of trials completed correctly in each of the four training conditions, plotted separately for the 2 subjects, Macduff and Oberon. No hints were available during Session 4. All trials, including trials on which the hint was available, were included in the analysis. The horizontal dashed lines represent baseline performance on a new list when hints were unavailable.

expect given their expertise at learning lists by trial and error (Terrace et al., 2003).

The absence of any learning on the nonhint trials under the 50%-hint condition could reflect the contribution of two non-mutually exclusive factors: one reflecting motivation and the other the effects of interference. Given that subjects could expect to earn a reward on 50% of the trials, they might not have been motivated to remember the consequences of responding on the nonhint trials. In addition, if subjects did not attend to the list items on hint trials in the 50%-hint condition, those trials may have served as interference, making whatever learning occurred on nonhint trials less likely to take root. In any case, performance under the 50%-hint condition shows that Macduff and Oberon did not always use hints in a constructive manner.

The fact that accuracy did not increase during the first 3 days of training in any of the three hint conditions is further evidence that hints did not foster learning (see Fig. 1). By contrast, accuracy increased steadily under the no-hint condition, in which subjects had to learn the lists by trial and error.

Performance When Hints Were Not Available

Subjects' performance during Session 4, when hints were not available, provided clear evidence that training without hints produced more long-term learning than training with hints. As shown in Figure 1, accuracy dropped precipitously in Session 4 for lists trained under the 100%-hint, 50%-hint, and auto-hint conditions, but not lists trained under the no-hint condition. On Day 4, there was a main effect of training condition on accuracy for both subjects, $F(3, 14) = 15.96$, $p < .0001$, $\eta_p^2 = .77$, for Macduff and $F(3, 16) = 21.11$, $p < .0001$, $\eta_p^2 = .80$, for Oberon. Post hoc Tukey tests comparing the four conditions showed that accuracy under the no-hint condition was significantly higher than accuracy under each of the other three conditions. None of the three hint conditions differed significantly from one another. Taken together, these findings indicate that the efficacy of training can be increased significantly by having subjects generate information from memory by trial and error.

Performance during Session 4 also provided evidence that training with hints did not produce any lasting learning. Accuracy in each hint condition dropped to the value expected with no prior training—that is, to the level of Session 1 performance in the no-hint condition (see the dashed lines in Fig. 1). Indeed, there was no significant difference in performance between Session 1 in the no-hint condition and Session 4 in any of the three hint conditions. By contrast, performance under the no-hint condition was significantly higher in Session 4 than it was in Session 1, $t(7) = 10.15$, $p < .0001$, for Macduff and $t(9) = 13.83$, $p < .0001$, for Oberon.

DISCUSSION

The enhancement of memory that resulted when monkeys generated responses during practice on serial lists is similar to the generation effect described in humans (e.g., Slamecka & Graf, 1978). The observed generation effect in monkeys is especially striking given that (a) subjects produced essentially the same behavioral responses during training with and without hints (locating and touching the list items in the correct order) and (b) hints produced no measurable learning, despite the high levels of performance during training.

Various explanations of the generation effect (and related testing effects—e.g., Roediger & Karpicke, 2006a) have suggested that, in the case of verbal learning, generation requires more attention, more effort, and more active information processing than memorizing the same information by passive observation (e.g., Hirshman & Bjork, 1988; Roediger & Karpicke, 2006b). Although active generation of answers during training may result in low initial performance, it enhances long-term retention and transfer (Bjork, 1994). In the current experiment, hints made the task easy—perhaps too easy—and bolstered immediate performance, but that gain was lost when the hints were removed.

The finding that the most effective training condition (no hints) exactly matched the testing condition is consistent with transfer-appropriate processing. Transfer-appropriate processing cannot explain performance during Sessions 1 through 3, however. During those sessions, there was little evidence of learning in any of the hint conditions: Performance did not increase across sessions, and performance on nonhint trials in the 50%-hint condition barely exceeded chance. Only no-hint training improved memory performance during Sessions 1 through 3. Thus, transfer-appropriate processing may have played a role in Session 4, but it cannot fully account for the current findings.

The way subjects processed information in the presence and absence of hints differed in the degree of generation; the nature of the information processed may have differed as well. At a basic level, subjects were required to attend to the photographs, in order to touch them, in all conditions. In the no-hint condition, however, they had to encode each item and its ordinal position in the list; they did not have to attend to ordinal position on hint trials. Knowledge of ordinal position is the only means of achieving high levels of performance in the absence of a hint.

When learning a sequence by trial and error, it is necessary to make some errors to learn which responses are not appropriate at different steps of the sequence. The need to make "logical" errors (Terrace et al., 2003) during the acquisition of a simultaneous chain provides an important clue as to why hints were not effective under any of the hint conditions in this experiment. A hint provides information about the next correct response, but not about errors.

Our results, which show that active learning facilitates nonverbal serial memory in primates, have broad implications for developmental psychology, neuroscience, and the evolution of intelligence. In particular, these results add to the growing list of nonverbal cognitive phenomena that can be studied in animals. These phenomena include numerical ability (Brannon & Terrace, 1998), serial expertise (Terrace et al., 2003), distance effects (Terrace, 2005), and metacognition (Son & Kornell, 2005).

Acknowledgments—This research was supported by a grant from the National Institute of Mental Health to H.S.T. (MH 40462). We thank Lisa K. Son and Jessica F. Cantlon for their assistance in this project and for their helpful comments about earlier drafts of this article.

REFERENCES

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Brannon, E.M., & Terrace, H.S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, *282*, 746–749.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642.

Glover, J.A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.

Hirshman, E.L., & Bjork, R.A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 484–494.

Hogan, R.M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562–567.

Jacoby, L.L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

James, W. (1890). *The principles of psychology*. New York: Holt.

Kornell, N., Son, L.K., & Terrace, H.S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, *18*, 64–71.

Roediger, H.L., III, & Karpicke, J.D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

Roediger, H.L., III, & Karpicke, J.D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Slamecka, N.J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.

Son, L.K., & Kornell, N. (2005). Metaconfidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In H.S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 296–320). New York: Oxford University Press.

Terrace, H.S. (2005). The simultaneous chain: A new approach to serial learning. *Trends in Cognitive Sciences*, *9*, 202–210.

Terrace, H.S., Son, L.K., & Brannon, E.M. (2003). Serial expertise of rhesus macaques. *Psychological Science*, *14*, 66–73.

(RECEIVED 11/6/06; REVISION ACCEPTED 1/16/07; FINAL MATERIALS RECEIVED 1/19/07)