Paper-Marking Test for Chimpanzee: Simple Control for Social Cues

Abstract. *A series of paper-marking tests, modeled after tests widely applied to humans, was administered to Sarah, a "language"-trained chimpanzee* (Pan troglodytes). *The test format was simple, economical, and controlled for social cues. The ape successfully made same-similar-different judgments on pictures of familiar objects when up to four questions were presented at the same time. Performance remained satisfactory on same-different judgments of novel pictures of alphabetic characters. Throughout the series of tests, the subject showed a superiority on same judgments.*

Development of refined procedures for testing animals has not kept pace with recent advances in comparative research on language and cognition. The several methods designed to control for social cues in testing "language"-trained animals prove too demanding for extensive application (*1*). Conventional animal-testing procedures furnish no better alternative. The widely used Wisconsin General Test Apparatus (*2*) does eliminate social cues, but is arduous, time-consuming, and hardly a method that would be used for comparative work with humans.

Tests designed for humans are simple and expedient, principally because they exploit the cognitive and linguistic capacities of the subject. The "paper-pencil" test one finds in a scholastic aptitude test is an illustration. This test allows presentation of multiple questions at the same time. The subject reads each question and leaves a mark on one of several alternative answers. If the instructor fears the intrusion of social cues, he may isolate the student, leave the room, and return later to grade the exam. This format is a powerful language-based test device: it invites application to the language-trained animal subject.

We designed a simplified "paper-pencil" test for chimpanzee which preserved the salient features of the human version. While this test procedure accommodates virtually any subject matter, we describe its first application in chimpanzee to relatively simple judgments of same, similar, and different.

The subject was Sarah, an African-born female chimpanzee (*Pan troglodytes*) approximately 14 years old. Sarah had lived in the laboratory for more than 13 years. During the last 10 years, she received five lessons per week on a number of cognitive tasks, including a simplified language (*3*). The words "same" and "different" had been taught to her previously. Like all of her words, these were pieces of plastic, metal-backed, which could be made to adhere to a magnetized slate. "Same" and "different" had been used as elements of plastic-word sentences written vertically on a slate, or in hybrid strings of both pictures and words written horizontally on a tabletop (for example, "apple" "same" 'apple,' where 'apple' was a photograph of an apple). In the original training, she was taught to replace one word (the interrogative particle, "?") with another (either "same" or "different"). She placed the appropriate word between pairs of like or unlike objects (for example, paper clip "same" paper clip, or paper clip "different" pencil). She eventually progressed to judgments of pairs of sentences on the basis of synonymy. For example, when presented with a plastic-word string such as "apple is red ? red color of apple," she replaced "?" with "same"; when presented with "apple is red ? cherry is red," she replaced "?" with "different" (*3*).

We made paper versions of the original object exemplar tests. Here, Sarah's task was to check or mark the one word that accurately described the relationship between two exemplars depicted on the test paper. Each question consisted of two exemplars and an interrogative particle, and was accompanied by two potential answers. A 3 by 4 cm picture of her original interrogative particle was centered on the top line of the page. The exemplars were 4 cm to its left and right. Answers were depicted on the bottom line of the paper, 7 cm apart. Five tests based on this format were administered in succession. The tests differed in the nature of the exemplars, the set of potential answers, and in the number of questions presented at the same time.

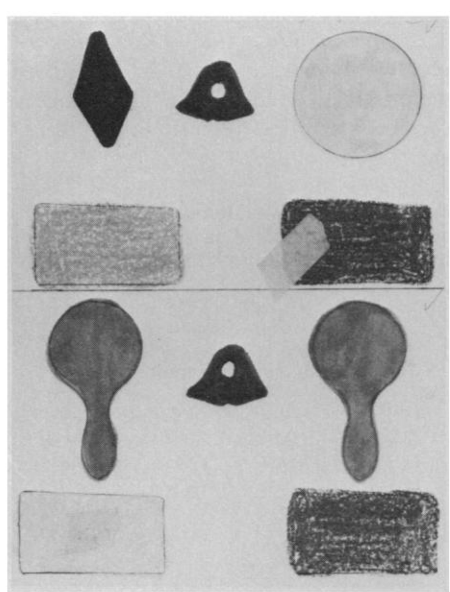

Fig. 1. An example of a test paper used in test 2.

In test 1, exemplars were pictures of familiar toy objects, generated from a set of 30 colors and 40 toys, ranging in size from 1 cm^2 to 6 by 9 cm. Pictures were made by tracing the outline of a toy on the paper and coloring in the form with crayon. A particular picture could appear more than once during the test, but each combination of two exemplars occurred only once during the entire experiment. Three potential answers were used, but only two appeared in the list accompanying any one question. Always one answer was correct and the other incorrect. Answers were pictures of plastic words, drawn in the same manner as that described for exemplars. Each word was a 4 by 7 cm rectangle, colored yellow for "same," orange for "similar," or red for "different." Exemplars that were labeled "same" were of the same shape, size, and color. Those labeled "similar" were identical on two of the three dimensions; and those labeled "different" were not identical on any of the three dimensions. "Similar" had not been taught previously, and the present test afforded the opportunity to assess Sarah's capacity to acquire a new word, as well as use her original words, in a pictorial mode. Each combination of two words from the set of three appeared approximately equally often during the test. Position of each word on the left or right of the page was counterbalanced across test papers. In this test, each page contained one question. Top (question) and bottom (answers) rows of pictures were drawn 7 cm apart on standard sheets of white typing paper (20 by 28 cm). One test paper was delivered to the subject on each trial, and was centered on a 36 by 46 cm plastic tray, for ease of presentation. Sessions consisted of 4 to 12 trials, separated by intertrial intervals of approximately 1 minute. Sessions were conducted once per day, three times a week.

In test 2, two questions were presented at the same time on every trial and both were drawn on one standard sheet of paper (Fig. 1). The paper was bisected by a heavy black line; one question appeared above the line, the other below. Space limitations required that top (question) and bottom (answers) rows on

0036-8075/78/1124-0903$00.50/0

each half of the page be drawn only 2 cm apart. Otherwise, the procedure was identical to that for test 1.

In test 3, four questions were presented on each trial. Questions were drawn as in test 2, but each sheet of paper was cut in half along the heavy black line. The four test papers were presented on two trays, two pages to a tray. Pages were 4 cm apart on each tray, and the trays were presented 60 cm apart at the beginning of each trial. All other aspects of the procedure remained the same as those for test 1.

In test 4, exemplars were letters of the English alphabet. Sarah had never been tested with letters in any previous task. This test allowed an assessment of her proficiency at judgments based on completely novel material. Letters were painted on the paper with black ink by means of a commercial stencil. In this test, the five letters A, B, C, Y, and Z formed the set of exemplars. Each letter appeared equally often, and position on the left or right of the top row was counterbalanced across test papers. Inasmuch as letters were always alike in color (black) and size (5 by 5 cm), only "same" and "different" were potential answers in the present test. As in previous tests, position of each answer was counterbalanced across successive test questions. We returned to the format of test 1 here, by drawing each question on a standard sheet of paper. One question was presented on the tray during each trial. No other aspects of the procedure differed from those for test 1.

Finally, in test 5, a new set of letters (D, E, H, I, and S) served as novel exemplars. Two questions were presented on each trial, one on each of two trays. Each question was drawn on a single standard sheet of paper, and each page was centered on one of two trays delivered 60 cm apart at the start of each trial. Otherwise, the procedure was the same as that for test 4. Figure 2 shows Sarah confronted by two problems of the form administered in this fifth and final test.

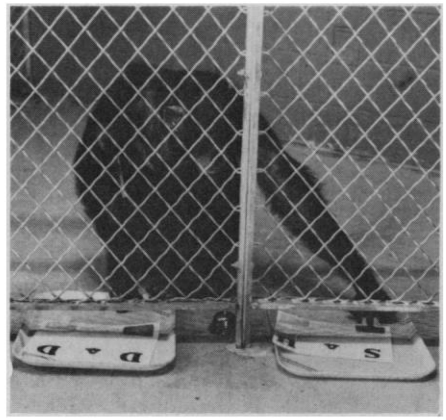

Fig. 2. Physical arrangement of test situation showing Sarah marking her answer with masking tape. The trainer, stationed in an adjacent room, is summoned by Sarah's use of bell.

Before the start of the test series, Sarah was taught to mark her answer with a pencil, with a combination of modeling and passive guidance techniques. Sarah cooperated at first, but her enthusiasm for scribbling both on and between pictures eventually overcame the strictures of the test. Although we could often interpret her answer as the word with the most prominent marks, we removed the pencil and adopted another approach midway through the second test. We gave her one piece of masking tape and in two demonstration trials trained her to put the tape on one of the words. Fortunately, she never devised any further innovations, such as tearing the tape in two and marking more than one answer. This system was used for the remainder of the experiment.

Twelve errorless trials (only the correct word appeared on the test paper) were used to adapt her to performing in the absence of the trainer. On each trial, the trainer entered the room containing her cage and pushed the tray holding one test paper under the wire mesh toward the subject. Sarah marked her answer and then rang a bell, signifying that she had completed the problem to her own satisfaction. (Her use of the bell was a common feature of prior tasks.) The trainer next graded the exam by pointing to her answer and telling her, "Good Sarah, that's right," or, "No Sarah, that's wrong," in a tone of voice one would use with a young child. On trials 1 to 5, the trainer remained by the cage while Sarah answered the question, coaxing her to respond if necessary. On trials 6 to 9, the trainer withdrew to the door of the room and watched until Sarah rang her bell. On trials 10 to 12, the trainer left the room, closed the door, and awaited summons from Sarah by means of the bell. This procedure was in effect for the remainder of the experiment during tests 1 to 5. At the end of every session, Sarah was given yogurt, fruit, or candy.

Despite the absence of the trainer, Sarah's performance on all five tests exceeded levels that would be expected by chance. Table 1 presents her overall score on each test, as well as scores for each type of judgment (same, similar, or different) in individual test formats. Her accuracy on same judgments remained quite high throughout the series of tests. In contrast, her accuracy on different judgments fluctuated from test to test, and failed to show a significant departure from chance in tests 1 and 3. Sarah's performance on the new case—similar judgments—began, as expected, at chance level in test 1. However, accuracy on "similar" increased across successive tests to a level comparable to that for same judgments. The overall pattern of results for tests 1 to 3 showed that Sarah experienced little difficulty in judging pictorial material, or in progressing from one to four questions per trial. Indeed, she rapidly acquired the new word "similar" during the course of testing. In the final tests, her performance with novel exemplars remained quite good. Her scores were above chance for both "same" and "different," although her superiority on "same" reappeared in test 5 (Table 1).

When results for the first three tests were combined, Sarah was most accurate on same judgments (88 percent correct), intermediate on similar judgments (76 percent correct), and least accurate on different judgments (70 percent). Her superior performance on "same" is in keeping not only with her own earlier

Table 1. Sarah's accuracy on same, similar, and different judgments during each test. Each score designates the number of correct answers divided by the number of questions of each type presented to the subject during each test. Probabilities were computed with the normal approximation to the binomial distribution (*7*).

Test	Sessions	Questions per trial	Nature of exemplars	Type of question			Overall score
				Same	Similar	Different	
1	3	1	Pictures of objects	11/13*	5/10	8/12	24/35*
2	6	2	Pictures of objects	23/24†	19/22†	18/24*	60/70†
2	3	4	Pictures of objects	17/21‡	11/14*	11/17	39/52†
4	4	1	Letters A, B, C, Y, and Z	16/20‡		16/20‡	32/40†
5	4	2	Letters D, E, H, I, and S	19/20†		15/20*	34/40†
Total				86/98†	35/46†	68/93†	189/237†

*$P < .05$. †$P < .001$. ‡$P < .01$.

performance with plastic words (*1, 3*), but also, in a sense, with human performance. Although accuracy in humans does not often vary (*4*)—and certainly would not on material of this simplicity—latency on same judgments is typically shorter than that on different judgments (*5*). We have been unable to find comparative data on similar judgments, however.

In tests 1 to 3, accuracy on "same" and "different" varied with the contrasted alternative. In questions in which "same" was the correct answer, Sarah was correct on 93 percent of trials when "different" was the alternative but only 82 percent when "similar" was the alternative. Similarly, in questions in which "different" was correct, she was correct on 80 percent when "same" was the other choice but only 60 percent when "similar" was the other choice. Thus, for both same and different judgments, clear-cut choices (same-different) met with greater accuracy than the finer distinctions (same-similar and similar-different). Also, in line with this conclusion were the results for similarity judgments, which did not vary with the contrasted alternative. In questions in which "similar" was correct, she was correct on 74 percent when the alternative was "same" and 78 percent when the alternative was "different."

The chimpanzee's success on the present tests demonstrated not only rudimentary cognitive and linguistic capacities but a sophisticated perceptual-motor capacity as well. The perceptually most notable feature of the tests, easily missed by the extraordinarily perceptive human observer, is that the question and its potential answers occurred on separate lines. The subject responded not by completing a gap in a string of objects (plastic words), but "read" the question depicted on the page, found the appropriate answer in a list of alternatives, and marked the answer, even though it was physically removed from the question. The results show that chimpanzee, like human, can do readily what many other species can do only with great difficulty, if at all: respond correctly to one location on the basis of stimuli presented elsewhere (*6*).

The present results cannot be interpreted in terms of the subject's sensitivity to inadvertent social cues. However, Sarah's performance may have relied critically on quite different features of her social relationship with the trainer. He was quick to respond to her summons at the end of every trial, and praise or food was soon to follow. Left to herself for longer periods of time with a test booklet, would she continue to respond? Would she answer only those questions on which she was proficient (for instance, only same judgments)? Indeed, does she know that some questions pose difficulties for her? Would she answer all questions rapidly, but review her answers later and make corrections? The potential of the present test format for providing answers to such questions can now be explored.

DAVID PREMACK
Department of Psychology, University of Pennsylvania, Philadelphia 19174

GUY WOODRUFF
KEITH KENNEL
University of Pennsylvania Primate Facility, Honey Brook 19344

References and Notes

1. D. Premack, *Science* **172**, 808 (1971); D. M. Rumbaugh, T. V. Gill, E. C. von Glasersfeld, *ibid.* **182**, 731 (1973).
2. H. F. Harlow and J. A. Bromer, *Psychol. Rec.* **19**, 434 (1938).
3. Details of procedures and results are given in D. Premack, *Intelligence in Ape and Man* (Erlbaum, Hillsdale, N.J., 1976).
4. Accuracy in humans occasionally does vary; some experiments show a bias toward "same," but just as many show the opposite effect. Investigators using human subjects most often report only latency data [E. A. C. Thomas, *Psychol. Rev.* **81**, 442 (1974); R. S. Nickerson, *Percept. Mot. Skills* **24**, 543 (1967)].
5. D. Bamber, *Percept. Psychophys.* **6**, 169 (1969); R. S. Nickerson, *Percept. Mot. Skills* **20**, 15 (1965).
6. D. R. Meyer, F. R. Treichler, P. M. Meyer, in *Behavior of Nonhuman Primates*, A. M. Schrier, H. F. Harlow, F. Stollnitz, Eds. (Academic Press, New York, 1965), p. 1.
7. W. L. Hays, *Statistics* (Holt, Rinehart & Winston, New York, 1963).
8. Supported by NSF grant BMS 75-19748 and by a facilities grant from the Grant Foundation. We thank A. J. Premack for helpful comments on the manuscript.

13 February 1978

Memory Impairment in Korsakoff's Psychosis: A Correlation with Brain Noradrenergic Activity

Abstract. *The concentration of the primary brain metabolite of norepinephrine is diminished in the lumbar spinal fluid of patients with Korsakoff's syndrome. The extent of its reduction is significantly correlated with measures of memory impairment for individual patients. These data suggest that the memory disorder of Korsakoff's syndrome may result from damage to ascending noradrenergic pathways by the diencephalic and brainstem lesions associated with this disease.*

Korsakoff's psychosis, the chronic phase of the Wernicke-Korsakoff syndrome, is an organic brain disease characterized by varying degrees of retrograde and anterograde amnesia with relative sparing of other intellectual functions (*1, 2*). The acute phase of the disease, Wernicke's syndrome, has long been attributed to a specific thiamine deficiency, typically occurring in nutritionally depleted chronic alcoholic individuals and generally improving after treatment with thiamine. Blass and Gibson (*3*) recently reported a defect in transketolase (E.C. 2.2.1.1), a thiamine-dependent enzyme, in cells cultured from patients with the Korsakoff syndrome; however, the relation between thiamine deficiency and the symptoms characteristic of Korsakoff's psychosis is not clear.

The neurologic basis of the memory deficit seen in Korsakoff's psychosis is not understood. Pathologic studies (*2, 4*) have demonstrated consistent diencephalic and brainstem lesions symmetrically located in the region of the third and fourth ventricles and aqueduct. It is of particular interest that these lesions are located along the pathways of monoamine-containing neurons that have been traced by histofluorescence methods (*5*).

If the memory disorder of Korsakoff's psychosis is related to impaired central monoamine systems, one might expect to find a change in the cerebrospinal fluid (CSF) concentrations of one or more of the brain metabolites of these putative neurotransmitters: namely, 3-methoxy 4-hydroxyphenyl glycol (MHPG) and vanillylmandelic acid (VMA), homovanillic acid (HVA), and 5-hydroxyindoleacetic acid (5-HIAA), central metabolic products of norepinephrine, dopamine, and serotonin, respectively. We have examined the CSF of nine patients with the Korsakoff syndrome and have found an abnormally low concentration of MHPG. We have also noted a correlation between the extent of decrease of MHPG in the CSF and the severity of memory impairment in individual patients.

A number of investigators have implicated ascending catecholamine pathways in learning and memory consolidation (*6, 7*). Anlezark *et al.* (*8*) found that mid-pontine lesions including locus coeruleus are associated with impaired subsequent learning and decreased cortical levels of norepinephrine in rats. Intracranial self-stimulation is supported by electrodes implanted in the locus coe-

 0036-8075/78/1124-0905$00.50/0