Computerized animal intelligence testing*

JOHN W. DAVENPORT, ROBERT W. BENSON, WILLIAM W. HAGQUIST
GENE R. RANKIN, and STEVEN E. SHELTON
Regional Primate Research Center
University of Wisconsin, Madison, Wisconsin 53706

In 1970, an on-line computer-controlled system for assessing the performance of monkeys and rats in batteries of learning tasks began operating in our laboratory. During the past year, we have used this system extensively for testing animals in mental retardation projects on hypothyroidism and malnutrition and have gained some appreciation of the system's utility for conventional learning studies of either the free-operant or discrete-trial type.

DESCRIPTION OF THE SYSTEM

The hardware of our system, summarized in Fig. 1, consists of a Digital Equipment Co. PDP-8/I computer which is connected to a group of monkey (or rat) test chambers through a Grason-Stadler SCAT 3001 interface and an additional relay interface of our own design. The computer is equipped with 8K 12-bit words of core memory, extended arithmetic, and a 32K disk memory for program storage. ASR-33 and RO-33 teletypewriters provide for control (TTY-IN) and data-output functions, respectively.

The interfacing provisions permit the control of up to five test chambers at a time, and, within each of the five "channels" emanating from the SCAT interface, there are lines for controlling up to eight stimulus devices (e.g., lamps, feeders, etc.) and for sensing up to three response manipulanda. A given channel may be time-shared by two different test chambers by means of the cable switcher, which consists of five multipole relays; thus, up to 10 chambers (e.g., five monkey and five rat chambers, as indicated in Fig. 1) can be controlled by the computer in a time-sharing mode.

The supplemental relay interface, which we constructed with 55 Potter and Brumfield KHP 17D11 relays, permits the handling of external devices (e.g., feeder solenoids) in the chambers having heavy current loads or voltages which must be isolated from the SCAT interface. A patch panel on the relay interface facilitates the use of auxiliary recording devices, such as cumulative recorders, and provides a simple means of routing stimulus lines to the few elements within multiple-lamp stimulus projectors (e.g., IEE display units), which are typically used within a given problem or test session. (For the types of learning problems we are currently studying, this manual feature obviates the purchase of additional stimulus lines beyond the eight per channel provided in the basic SCAT hardware.) A lamp panel (MONITOR) connected to the relay interface provides a visual readout of the stimulus and response events in each test chamber.

The monkey test chambers, which we call blockhouses, consist of homemade enclosures constructed of 4-in. solid concrete-block walls. Insulated plywood front doors and ceilings having observation windows complete the enclosure of the cubicles. Figure 2 shows three of the blockhouses with their front doors removed, revealing the interior stainless steel compartments, which are also removable for cleaning and for apparatus alterations. The interior dimensions of the compartments (46.0 cm wide, 60.5 cm deep, 64.0 cm high) are designed for infant and adolescent rhesus monkeys. Within a typical compartment, the monkey is confronted with a response panel equipped with stimulus-display pushbuttons, three-color illuminated retractable levers, and food-reward receptacles for a variety of learning tasks. The chambers are variously equipped with Davis universal feeders, liquid dispensers, air-blast pipes, or frosted screens (for projection of color slides or films) to provide a choice among food, milk, aversive, and visual reinforcers, respectively. Lehigh Valley test chambers for macaques and squirrel monkeys are also available for tasks involving the use of grid shock.

The rat test chambers are Gerbrands Model C units, enclosed in picnic ice chests. Each of the response panels in these chambers has left- and right-mounted Lehigh Valley retractable levers (which, like the monkey chambers, have frosted Plexiglas windows immediately above the lever paddles to provide illumination), a center-mounted pigeon key, a food receptacle (beneath the key), and a Davis pellet dispenser. Cables and connectors of the rat chambers and other test apparatus in the laboratory (e.g., automated runways and Y-mazes) are similar to those for the monkey blockhouses and Lehigh Valley chambers in order to provide maximum flexibility in the use of the computer for rat and monkey testing.

Events, contingencies, and recording provisions within a test session are governed by the E's software program, which is initially written in the SCAT (State Change Algorithm Terminology) language. The programming requirements for nearly any familiar task in animal learning can readily be converted into a sequence of instructions in this easily learned language. Further convenience is provided by Grason-Stadler's software packages, the editor/compiler, the assembler, and the operating system, which are stored for ready access on the disk. The editor/compiler allows the E to enter his SCAT-language program into the ASR-33 keyboard and provides error

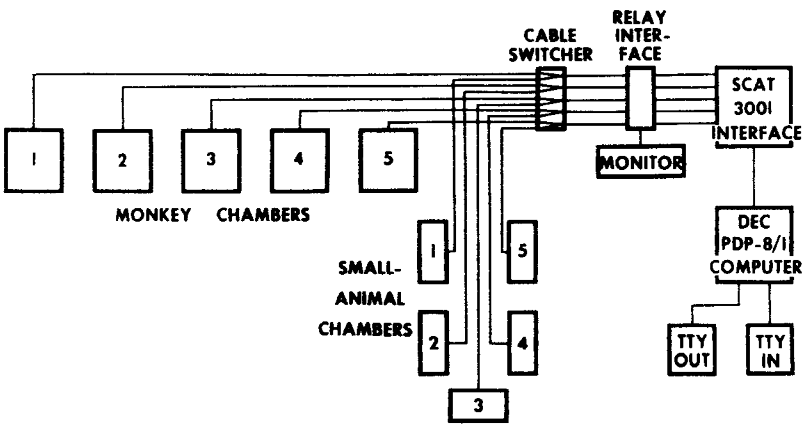

Fig. 1. Relationships among the major components of the computer-controlled testing system.

*Publication 11-017 of the Wisconsin Regional Primate Research Center. The facilities described were funded from the Center's base operating grant, RR-00167, awarded by the Division of Research Resources, National Institutes of Health. We thank Will Retzlaff and Bruce Preslan for their valuable contributions in the design and use of the testing systems and Susan Engelke for her expert assistance in processing data on the UNIVAC 1108.

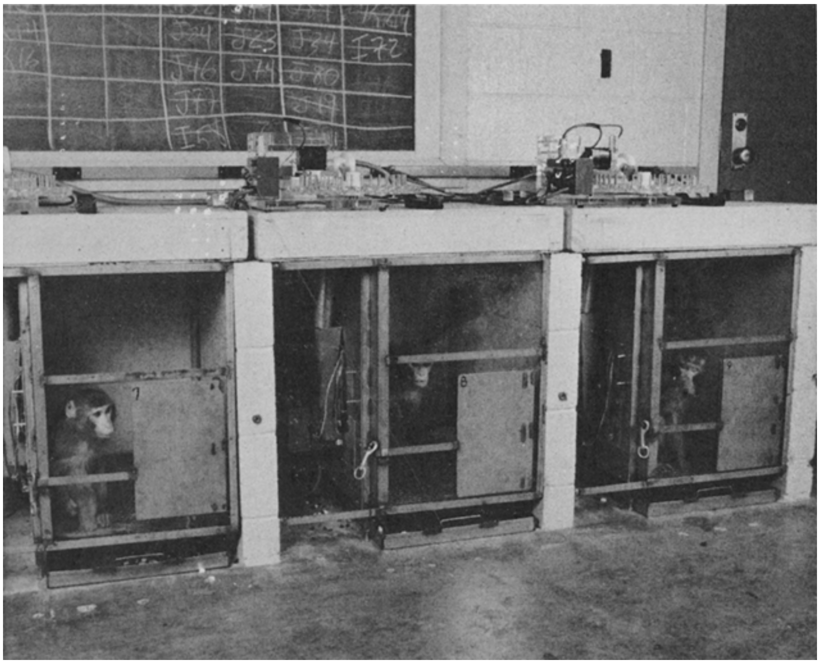

Fig. 2. Three of the "blockhouse" test chambers which are controlled by the PDP-8/I computer and SCAT interface. Each testing compartment is formed by two concrete-block walls, a response panel, a Plexiglas front, mesh flooring, and a rubber-lined plywood ceiling.

checking and diagnostics; this package also outputs an intermediate-language tape from which the assembler produces a binary tape acceptable to the operating system. After the operating system has been read into core memory, the binary-tape program is fed into the ASR-33 tape readers, once for each remote station to be run on that program, and the system is ready to run. Future uses of the same program require only the last step; thus, as soon as the E accumulates a small library of binary tapes representing translations of debugged SCAT-language programs, he is in a position to change from one learning task to another with facility (see recent Grason-Stadler literature for further details about software).

Within the limits of the SCAT functionals (e.g., 10-msec clocks, 20 data-storage elements per station, two variable-number generators, various counting, comparing, arithmetic, and logic functions, essentially all that can be done with electromechanical or solid-state programming modules is readily achieved through the SCAT software. As with any up-to-date computer-controlled system for process control and data acquisition, of course, certain programming contingencies, recording features, and data-reduction functions are possible with the PDP-8/I-SCAT system which would be difficult or impossible to achieve with the conventional modules. The advantages of the former are particularly clear in situations requiring very rapid or complex calculations in the midst of a trial or time segment in order for a schedule contingency to be implemented, in experiments requiring accurate monitoring of events in frequencies approaching 100/sec, and where there is need for immediate outputting of summary statistics (e.g., histograms, means of response measures for various intermixed types of trials within a session).

For simple learning paradigms, such as free-operant and discrete-trial conditioning tasks involving a single response measure and sessions of customary length, there appear to be ample provisions in our system to permit a great variety of studies to be conducted in five chambers at a time. Es who wish to use a system such as ours for relatively complex learning studies, however, may end up sharing a major disappointment with us. This stems from the fact that, of the 8K total core-memory space, approximately 5K is required for the SCAT operating system package, leaving only around 3K for user programs controlling remote test chambers. Even with modest recording requirements (e.g., one latency measure and one choice indication per trial in a 60-trial session), a user program can readily exceed 20% of the available core space, in which case all five chambers cannot be run on that program simultaneously. To make matters worse, even the simplest learning paradigms will require large proportions of the available core space if multiple response measures are employed, particularly if the RO-33 is used as a printout counter for individual trial measures in long sessions. We have frequently been forced to cut back to only three chambers at a time, with efficient programs for discrete-trial discriminations (e.g., two-choice visual discrimination, a sameness-difference problem, matching to sample) and some free-operant paradigms (progressive ratio, multiple FR/DRL schedules). The only remedy for this problem seems to be expansion of core by another 4K.

USES OF THE SYSTEM

Our primary use of the system is for comparisons of normal and "retardate" animals on batteries of learning tasks. We refer to this activity, informally, as intelligence testing, since our test batteries presumably sample a wider range of learning capacities than any single task would, and our major aim is to compare empirical ability profiles for diverse groups of animals. In the first 3 years of this work, it has become quite apparent that the achievement of "final-draft" intelligence batteries whose component tasks are empirically validated is at least a 10-year proposition. Besides developing specific procedures for administering each task and determining the practical feasibility of multiple-task batteries, the component tasks must be examined in terms of their sensitivity in detecting normal-retardate differences with some consistency (e.g., over several experiments). It may well prove to be the case that a majority of familiar nonsocial learning tasks for monkeys and rats lack such sensitivity.

Using conventional relay equipment as well as our computer, we have found it feasible to administer unusually long test batteries to rats within a reasonable period of time. Our recent experience includes the administration of five- and seven-task batteries to normal and hypothyroid rats within 1-year periods (Davenport, 1970), a nine-task study of thiouracil-treated rats which required 15 months for the testing, and an intercorrelational study with normal rats involving 22 tasks given over 13 months.

The tasks for rats have been conducted in automated runways, Y-mazes, shuttleboxes, and complex closed-field mazes, as well as in the operant chambers described earlier, but lately we have come to prefer discrete-trial barpressing analogues of the runway for such tasks as successive acquisitions and extinctions and two-lever analogues of the Y- and T-maze for spatial discrimination, successive reversals, nonspatial discrimination (between flashing and steadily illuminated levers), sign-differentiated discrimination (left or right lever correct, depending on concurrent tone frequencies), and short-term memory (extension of the latter, with delays interposed between tone and lever presentations). In our current intelligence testing of rats, we are using batteries in which these analogue tasks are supplemented by motivational tests (e.g., running-wheel activity, rate of free-operant barpressing) and by testing in manually scored symmetrical maze problems (Davenport, Hagquist, & Rankin, 1970). Thus far, sufficient evidence of sensitivity to normal-hypothyroid differences has appeared in our data from the symmetrical maze and some of the leverpressing tasks (namely, spatial, nonspatial, and sign-differentiated discriminations) to justify their continued use in test batteries for rats.

In our work with monkeys, we are examining automated forms of some familiar primate learning tasks, including visual discrimination reversal, learning set, matching to sample, and short-term memory (delayed matching), primarily because the animal-learning and mental-retardation literatures provide advance indications that these tasks have the capacity to distinguish among species and age levels as well as between normal and retardate monkey groups. Supplementing these are some tasks which have shown sensitivity in detecting deficits in thyroidectomized or malnourished monkeys, including successive acquisitions and extinctions, discrimination between correlated and uncorrelated reinforcement, and discriminative avoidance.

Some formidable barriers exist in the administration of multiple-task batteries to monkeys which, by comparison, are minimal in similar testing of rats. Among these are the much higher cost of maintaining primates, the continual threat of interruptions in testing because of illness in the colony, the extended periods of time required for adapting the monkeys to the test apparatus, and the difficulty in maintaining constant motivational levels over the course of testing. We have embarked on our intelligence research with monkeys, therefore, with resignation to the fact that testing can proceed at only about half the pace of rat testing.

In two current studies with rhesus monkeys, we have spent several weeks on the pretraining phase, frequently scheduling 24- or 48-h sessions on weekends for monkeys which were unusually slow in adapting to the blockhouse chambers. Only three out of 61 monkeys have been dropped because of failure to adapt. The pretraining procedures include magazine training, lever shaping, incentive preference tests, and a "curiosity test" in which reward for leverpressing is shifted from food to color-slide presentations. The software programs for these procedures require relatively small amounts of core-memory space, and, thus, running five monkeys at a time, or from 20 to 40 monkeys within the normal working hours of a day, has been feasible in this initial stage of testing.

The programs for most tasks beyond pretraining require more than 20% of the available core space, however, and force a restriction to three or four chambers at a time. This restriction is somewhat offset by the fact that the SCAT interface can control up to five different tasks simultaneously. Thus, if the program for two-choice visual discrimination takes up 26% of the available core (permitting only three chambers to run on that problem), the remaining core space (22%) can be used during the same time for, say, magazine training (8%) in the fourth chamber and lever shaping (12%) in the fifth chamber. This arrangement is common and particularly suitable to our research, since we must hold constant the age at which a given task is administered to an animal and the monkeys within a given study are usually of disparate ages.

Some more detailed advantages of the system are illustrated by our task involving discrimination between correlated and uncorrelated reinforcement. In this paradigm, the monkey receives discrete trials in which he is confronted with a "DRL" lever (reward for pressing only if latency exceeds 5 sec) and a second lever whose payoff is independent of the response latency to it but which is yoked to the number and sequence of rewards successfully attained on the first (correlated) lever. A programmed sequence of "forced" (one-lever) and "choice" (two-lever) trials is given, with lever presentations contingent upon a trial-initiating pushbutton response on all trials. In each session, there are five seven-trial blocks; the first six trials of each block are a random intermixture of three correlated- and three uncorrelated-lever forced trials. On the seventh trial in each block, both levers are presented, giving the monkey a choice; the 5-sec DRL contingency prevails for the correlated lever as usual on this choice trial, and reward for choice of the uncorrelated lever is sometimes received or sometimes not, depending upon the animal's level of success on the correlated lever in the immediately preceding forced trials (reward is "set up" if there were two or three successes, but not if success was below two out of the three possible successes). Normal animals eventually learn to prefer the uncorrelated lever on choice trials and to respond faster to that lever on forced trials, presumably because, although reward is equally frequent on the two levers, the animal can get it sooner when it is available on uncorrelated-lever presses.

Besides providing for appropriate intermixtures of forced, choice, correlated-, and uncorrelated-lever trials and the various intra- and intertrial contingencies, the program for this task also generates an output of speed measures (reciprocalized latencies) for each of the 35 trials in a session, mean speeds for correlated- and uncorrelated-lever forced trials in each block and over all blocks, and number of correct (uncorrelated-lever) choices in the five choice trials. This program utilizes 32.6% of the available core space for each of three test stations. In testing animals on this and other tasks, we have come to feel some delight in the computer system's advantages over relay technology in nearly every aspect of the control and recording procedures.

We are getting additional mileage out of the PDP-8/I, during hours when it is free from on-line use, for statistical summarizing of data on punched tape from our relay-controlled experiments. The major portion of the data analysis in our intelligence research, however, is done on a UNIVAC 1108 elsewhere on campus after summary measures from our output TTY are manually punched onto cards. The 1108 programs provide analyses of variance for group comparisons, reliability indices, task intercorrelations, cluster analyses, and convenient methods for converting raw scores to standard scores and combining scores over an entire battery for composite "IQ" measures.

The provisions which now define Grason-Stadler's basic SCAT system (without options) are essentially identical to those in our on-line system. Although we feel continually pressed for space—in core-memory size and in the number of stimulus and response lines per station—the system is running well (down-time this year averaging only about 1 day every 2 months) and is in general delivering the promises of the company's most recent advertising. It suits our research purposes rather nicely, despite the fact that the original SCAT designers probably had conventional operant conditioning research in mind, and it seems likely to exceed our initial expectations as we become more skilled in realizing the full potential of an on-line computer.

REFERENCES

DAVENPORT, J. W. Cretinism in rats: Enduring behavioral deficit induced by tricyanoaminopropene. Science, 1970, 167, 1007-1009.

DAVENPORT, J. W., HAGQUIST, W. W., & RANKIN, G. R. The symmetrical maze: An automated closed-field test series for rats. Behavior Research Methods & Instrumentation, 1970, 2, 112-118.

Using the PDP-12 in verbal learning and short-term memory research*

BENNET B. MURDOCK, PHILIP O. DUFTY, and RONALD OKADA
University of Toronto, Toronto, Ontario, Canada

In this paper, we would like to do four things: (1) discuss the rationale for using an on-line computer to conduct research in the verbal learning and short-term memory area; (2) describe our operating system; (3) list the experiments currently in progress or planned; (4) evaluate our progress to date. Of necessity, all sections will be brief.

RATIONALE

An issue of some general concern is whether an on-line computer is a necessary laboratory device for the experimental psychologist. Prospective purchasers and funding committees often weigh heavily the issue of whether the research is of such a nature as to be impossible without a computer. A related question, often asked by those fortunate (or unfortunate) enough to have a computer in their laboratory, is whether they are making full use of its capabilities. Given the power and flexibility of a computer, is the psychologist really exploiting it to the full?

In our opinion, this is the wrong question to ask. It loses sight of the purpose of psychological experimentation. The purpose of experiments is to test theories and extend our empirical knowledge. The question that should be asked by the prospective or reflective user is whether he will do more and/or better experiments with a computer than without a computer. The purpose of doing experiments is *not* to utilize our equipment; it is to expand our knowledge.

As a simple example, if one has a multifactorial analysis of variance to do, one can spend hours or days at a desk calculator doing it by hand. Alternatively, one can run an ANOVA program on the university computer. In so doing, one is using a multimillion-dollar device as a simple adding machine. Considering matters of time and efficiency, there is little doubt which alternative most psychologists would prefer.

Further, while everything else is going up in price, on-line or "mini" computers are coming down, dramatically. The lowest-priced computers are not even necessarily the most expensive device one might have in a psychological laboratory, and even medium-priced laboratory computers are comparable to the salary one would pay a good technician over a period of several years. However, software costs are something else again, in terms of either trained personnel and/or the investigator's time required.

OPERATING SYSTEM

We have developed an operating system (OS12) to run experiments in verbal learning and memory using a PDP-12A (Dufty, 1970). Originally, our machine had only 4K of core, and the operating system requires about half of it. We have expanded to 8K of core to use FOCAL-12, but have not yet finished modifying the operating system. The general purpose of OS12 is to facilitate the writing of programs to run experiments. It is in modular form so that additions, deletions, and modifications will be relatively easy. The ultimate goal is to write a program using only call statements to subroutines, but at present the programs must be written using some assembly language in addition to OS12.

The four principal features of OS12 are the use of FORTRAN-like input/output statements, nested interrupt levels, debugging aids, and provision for chaining and overlays. The read statement is of the general form, READ;DEV;DATA, where DEV is the input device and DATA specifies a starting location. The input devices may be core, the Teletype, buffered Teletype, LINC tapes, sense switches, keyboards (not yet operational), or the A/D converter. For output, the general statement is WRITE;DEV;DATA, where the devices may be core, the Teletype, buffered Teletype, display (the CRT), LINC tapes, relays, or the speaker. A typical format statement might be TEXT %(4A2"—"4A2)%, where 4A2 permits up to eight alphabetic characters to be displayed and the quotation marks are for literals. If the device for this format statement was the display, then the paired associate, GALLANT—LEGEND, might be shown on the CRT. Provided formats are alphabetic, binary, octal, decimal, and free.

For the interrupt, at the higher level, a counter is incremented to keep track of real time and to determine when the lower level is to be entered. At the lower level, the display is refreshed and other device routines are entered. The clock is the only device that can cause program interrupts. We have our own hardware clock with a normal frequency of 200 Hz. The slow interrupt is entered every 25 msec and the fast interrupt every 5 msec. The software can handle clock rates up to 5,000 Hz.

OS12 has a small monitor which is useful in debugging. It permits the command G to start a program, C to continue from a STOP (judiciously inserted at various points in a program), O to load into core ODT (Octal Debugging Technique, a standard software package), S to start ODT, D to return to DIAL (Display Interactive Assembly Language, an excellent editor and assembler particularly suited for the PDP-12),

*Supported by Grants APA 146 from the National Research Council of Canada and OMHF 164 from the Ontario Mental Health Foundation. The first author is responsible for writing this paper and the opinions contained herein, the second author for the operating system, and the third author for writing some of the programs for the described experiments.