Controlling experiments with a general-purpose time-sharing system

L. ROWELL HUESMANN, YALE UNIVERSITY, New Haven, Connecticut 06510

For many experiments in the behavioral and biomedical sciences, there is no substitute for an experimental-control computer with true real-time capabilities. However, if an experimenter can tolerate timing errors of several seconds, then he can reap the benefits of computer control by buying time on a general-purpose time-sharing system. There can be little doubt that a self-contained computer-controlled laboratory will be more economical if a fair number of experiments are contemplated. But when only a few experiments are planned, the researcher may find that he can run them with a general-purpose time-sharing system at only a fraction of the cost of establishing a computer-controlled laboratory.

Computer-controlled experimentation is becoming more and more popular in the behavioral sciences. The increasing need for precise control of independent variables and rapid accumulation of large quantities of data has engendered the widespread use of small computers to control experiments. The cost of small computers has decreased considerably in the past 10 years, making the use of such computers financially feasible. Yet, the cost is still substantial enough that many Es who could derive benefits from computer control are unable to afford it. This paper suggests an alternative approach for those human performance researchers who wish to reap the benefits of computer control but are reluctant to establish a complete computer laboratory.

DESIRABILITY OF COMPUTER CONTROL

An experimental-control computer not only permits many experiments that would otherwise be impossible, but also performs most experiments more efficiently and with less chance of error than a human investigator could. The advantages of computer control have been discussed in detail elsewhere (Uttal, 1968) and will only be reviewed here. Basically, a computer has four inherent advantages over the human E. It has a large infallible memory in which data can be stored and from which data can be retrieved almost instantaneously; it can perform simple operations very rapidly; it will execute any algorithm in exactly the same manner time and again; and it will never become tired or bored. As a result, more variables can be controlled in a computerized experiment and more data can be collected in a short time. Furthermore, the variables can be controlled more precisely. Contrary to the claims of some Es, no two Ss ever receive "exactly" the same treatment, but with a computer one can approach much closer to the ideal. However, many researchers are attracted to computer control by more specific advantages. Stimuli can be presented according to very complex rules based on a S's previous responses without causing any delay in the experiment. For example, Johnson (1969) controls a categorization experiment with a computer so he can maximize the amount of time a S needs to find the correct classification concept. After a S classifies an object, the computer calculates which answer (right or wrong) would eliminate the fewest possible concepts and provides the S with that answer. The correct concept is the one remaining after all others have been eliminated. To some researchers the experimental-control computer's most useful attribute is its ability to time accurately

many phenomena at once. To others, it is its ability to store large amounts of data without human intervention. With a computer, it is possible to convert a S's push-button responses into a statistical test of a hypothesis without a human touching or even seeing the data. Obviously, conscientious Es will want to study the raw data as well, but by automating this analysis process, the probability of data-processing errors is greatly reduced. Finally, the very procedure of constructing a program to conduct an experiment is useful since it forces the investigator to specify every aspect of the experiment unambiguously.

COSTS OF SMALL EXPERIMENTAL CONTROL COMPUTERS

Although several small computer manufacturers offer digital computers for about \$10,000, it is unrealistic to expect a "general-purpose" computer-controlled laboratory to cost less than five times that amount. Two sample configurations actually in use are diagrammed in Figs. 1 and 2. Both of these configurations cost over \$50,000 and are designed to run several Ss simultaneously in one of a few experiments at the same time.

The most costly addition to a small computer is an expanded memory. Yet the memory provided with the "basic price" is often, perhaps even usually, inadequate for the proposed applications. The next most expensive items are mechanisms for rapidly transferring data in and out of storage and for the long-term storage of programs and data. Most inexpensive configurations are constructed around a teletype console and a paper-tape reader and punch. Yet, even very high-speed paper-tape readers and punches prove cumbersome and frustrating for seemingly simple applications. Reading a program into the computer can be a major undertaking requiring several minutes, while translating and debugging on paper tape is inefficient and even more time consuming. Outputting data on paper tape is also time consuming and may interfere with the experiment's timing if large quantities are being collected. One is faced with this problem of rapidly outputting data since most of the "basic configurations" for experimental control are inadequate for sophisticated analysis. Hence, one can either buy additional hardware to upgrade the configuration or transfer the data to another computer for analysis. For all of these reasons, the E wishing to employ a computer to conduct experiments usually finds that high-speed input/output devices and auxiliary mass storage are a necessity. If storing programs for rapid retrieval and outputting experimental data are all that is required, a magnetic-tape unit may prove sufficient. However, if the investigator wishes to run several Ss at once, or more than one experiment at a time, or if he has a high rate of data acquisition (e.g., heart rate), he will find he needs a more expensive higher-speed random-access mass-storage device, such as a disk or a drum. Even when not absolutely necessary, a drum or disk greatly improves the operation of an experimental-control computer.

The addition of extra memory, high-speed input/output devices, and auxiliary mass-storage devices alone can easily drive the cost of a \$10,000 system to over \$50,000. However, there are other costs as well. If the Ss are to respond on teletypes or special devices, the cost of these must be considered. Furthermore, experience has shown that even a simple system profits from the attention of at least a part-time systems programmer and an electronics technician. Most current small computers perform

Fig. 1. A 12-bit computer system used for experimental control at the Thurstone Psychometric Laboratory, University of North Carolina.

Fig. 2. A 16-bit computer system used for experimental control in the Psychology Department, Carnegie-Mellon University.

with high reliability considering their complexity; yet malfunctions do occur and the user without a maintenance contract or technical assistance can be lost.

But the above considerations need not drive human-performance investigators away from computer control. New experiments can be automated by writing programs instead of buying new equipment. Variables can be controlled precisely, and responses can be measured and timed with remarkable accuracy. Large amounts of data can be collected rapidly and processed without error-inducing human handling. For \$50,000 to \$100,000, a computer-controlled laboratory can be constructed that will provide efficient operation of almost any type of experiment. For \$10,000 to \$50,000, a system adequate for a very specific application can be purchased. The question is: What can those researchers do who cannot afford such a system or who need it for only one experiment?

THE QUESTION OF RESPONSE TIME

All experiments are "real-time" studies in the sense that there are timing bounds within which certain events must take place. However, there are wide variations in just how accurately experiments must be timed. For some researchers, delays or timing inaccuracies of several minutes are acceptable; for others delays of a few milliseconds are unacceptable. Most perception experiments probably fall into this latter class, while most studies of complex problem solving fall into the former. For the purpose of this discussion, a division will be made between those experiments that require timing accurate to *less than 5 sec* and those experiments that do not.

The central theme of this paper is that many researchers who require timing accuracies no shorter than several seconds can conduct their studies on large general-purpose time-sharing systems designed primarily for mathematical data processing. This is not feasible at present for those who need more accurate timing. But advances in the hardware and software of large time-sharing systems may open this possibility (at least to 0.1 sec accuracy) in a few years. At present, probably the best solution for an E who needs millisecond accuracies but can't justify \$50,000 is to build a very specific system for the particular application he has in mind. By resisting generality, he may

establish a relatively inexpensive computer-control system for his needs. In the long run, abandoning generality for cost usually turns out to be a poor strategy; so an E should investigate thoroughly the possibility of establishing a general-purpose laboratory with other investigators before adopting this approach.

GENERAL-PURPOSE TIME-SHARING AND EXPERIMENTAL-CONTROL PROGRAMS

Introduction

Large general-purpose time-sharing systems are usually designed for data-processing applications. A time-sharing system (see Green & Huesmann, 1967) is a computer system that allocates very small quanta of time to each user in turn (e.g., 100 msec). When a user exceeds his quanta or pauses for a "slow" input/output operation, another user begins executing. On good systems it appears to each user that he is the only person on the computer. Typically, these systems have response delays varying from milliseconds to several seconds, depending upon the user's request. Requests that require accessing auxiliary storage may initiate delays longer than 5 sec. The response time of such a system is also quite dependent on the scheduling algorithm in the computer's supervisory control program (monitor). Given a reasonable set of constraints upon such a scheduling algorithm, a user of a large time-sharing system can expect his program to have a response time of well under 5 sec. Of course, requests to access personal files in auxiliary storage may produce longer delays. The important point is that users of a large time-sharing system will usually have enough storage space available so that they will not need to access auxiliary storage. Hence, it is feasible to run real-time experimental-control programs off a general-purpose time-sharing system as long as 2- or 3-sec delays can be accepted. For this to be financially feasible, the charges on the time-sharing system must be based primarily on actual computing time and not on connect time. Finally, the user must be able to access the computer's real-time clock.

If these requirements are satisfied, the E can reap the advantages of computer control by using the general-purpose time-sharing system. In fact, there are some ways in which such a system is preferable to a small computer. Most experimental-control programs deal extensively with character

and string manipulations. Although the programs could be written in most any language, they would be much easier to write in a language such as PL/1, LISP, or SNOBOL. Such languages are seldom available on small computers, but frequently available on large general-purpose systems. Equally advantageous may be the immediate availability of the powerful data-processing routines found on general-purpose systems. Sophisticated data analyses on large amounts of data can be carried out while the experiment is proceeding. Data need not be transferred to a larger machine for processing.

Ideally, one would like several additional characteristics in the time-sharing system though they are by no means necessary. It would be nice if a user could specify routines to execute on an interrupt basis. In other words, the routine only executes when some event occurs, e.g., an input. Until recently, such a facility would not have been dreamed of, but some recent languages for time-sharing systems do include such facilities, e.g., PL/1. For many experiments, a time-limit interrupt of this type would be a necessity. It would also be nice if the S could be run on devices other than a teletype. Many experiments can best be accomplished with CRT display scopes on which data can be presented rapidly in verbal or graphical form. Still other investigators might need special-purpose devices. Attaching these would usually require special lines to the computing center.

Obviously, large areas of experimentation are excluded by restrictions inherent in this technique. However, many human-performance experiments can be carried out quite adequately with such an approach. Let me now describe three experiments that are being carried out on Yale's general-purpose time-sharing system designed for data processing not real-time control.

Example 1

One of the difficulties in studying learning strategies in competitive situations is that the S's behavior is intimately related to his opponent's, which in general cannot be held constant. This experiment was designed¹ to study a S's learning strategies while playing a game against a computer.

A simple, finite, two-outcome game was selected. The program was designed to play the game in either a learning mode, where it begins as a naive player, or in an expert mode, where it plays with the optimal strategy. A S may be instructed to win as many games as possible or only to try to derive the optimal strategy.

If the S plays against the optimal-strategy program, the program's moves will always be exactly the same in the same situation. If he plays against the learning version, the program's moves will be a function of previous moves and a random-number generator. The learning scheme is a simple "tree-pruning" scheme described by Block (1965).

The entire sequence of moves and countermoves is coded and automatically stored on the disk for analysis. The S's response latencies (accurate to within a few seconds) are also recorded. The time-consuming and often error-inducing process of transcribing the data and then inputting it into the computer is avoided. Sophisticated data analysis can be carried out immediately after the S is run.

A sample game between a S and the learning version of the program is shown in Fig. 3.

Example 2

This experiment was designed to be a variation of the Aronson and Carlsmith (1962) experiment on self-consistent behavior. They had discovered that under certain circumstances Ss who expected to do badly on a task and did well would change their behavior to do badly. On the other hand, Ss who usually performed well and suddenly performed badly would not change

YOU ARE GOING TO PLAY A GAME AGAINST A COMPUTER. THE COMPUTER WILL BE LEARNING TO PLAY THE GAME JUST AS YOU ARE. YOU SHOULD TRY TO LEARN THE OPTIMAL STRATEGY FOR PLAYING THE GAME. IF YOU FOLLOW THAT STRATEGY, THE COMPUTER WILL NOT BE ABLE TO BEAT YOU. HOWEVER, THE COMPUTER IS ALSO TRYING TO LEARN THE OPTIMAL STRATEGY. YOUR OBJECTIVE SHOULD BE TO WIN AS MANY GAMES AS POSSIBLE AGAINST THE COMPUTER.

EACH GAME BEGINS WHEN YOU TYPE A NUMBER BETWEEN 9 AND 19. THIS NUMBER WILL BE CALLED "NUM." THEN YOU ALTERNATE MOVES WITH THE COMPUTER (THE COMPUTER MOVES FIRST). ON HIS MOVE EACH PLAYER SUBTRACTS 1, 2, OR 3 FROM THE EXISTING VALUE OF "NUM." THE OBJECT OF THE GAME IS TO LEAVE YOUR OPPONENT WITH NUM=1.

ENTER INITIAL VALUE OF NUM.

>9
THE INITIAL VALUE OF NUM IS: 9
MY MOVE IS TO SUBTRACT: 3
THE PRESENT VALUE OF NUM IS: 6
YOUR MOVE:

>1
THE PRESENT VALUE OF NUM IS: 5
MY MOVE IS TO SUBTRACT: 1
THE PRESENT VALUE OF NUM IS: 4
YOUR MOVE:

>3
THE PRESENT VALUE OF NUM IS: 1
MY MOVE IS TO SUBTRACT: 1
THE PRESENT VALUE OF NUM IS: 0
WELL PLAYED, YOU WIN!!!

SINCE LAST INITIALIZATION, THE TOTAL WON-LOSS SCORE IS:
PLAYERS: 1 COMPUTER: 0
IF YOU WANT TO PLAY AGAIN ENTER "1" WHEN CUED. IF NOT ENTER "2."

>1
>14
THE INITIAL VALUE OF NUM IS: 14
MY MOVE IS TO SUBTRACT: 1
THE PRESENT VALUE OF NUM IS: 13
YOUR MOVE:

>4
THE PRESENT VALUE OF NUM IS: 9
MY MOVE IS TO SUBTRACT: 1
THE PRESENT VALUE OF NUM IS: 8
YOUR MOVE:

>1
THE PRESENT VALUE OF NUM IS: 7
MY MOVE IS TO SUBTRACT: 1
THE PRESENT VALUE OF NUM IS: 6
YOUR MOVE:

>6
YOUR MOVE MUST BE EITHER 1, 2, OR 3
REENTER MOVE WHEN CUED.

>3
YOUR MOVE
THE PRESENT VALUE OF NUM IS: 3
MY MOVE IS TO SUBTRACT: 2
THE PRESENT VALUE OF NUM IS: 1
I WIN!

SINCE LAST INITIALIZATION, THE TOTAL WON-LOSS SCORE IS:
PLAYERS: 1 COMPUTER: 1
IF YOU WANT TO PLAY AGAIN ENTER "1" WHEN CUED. IF NOT ENTER "2."

Fig. 3. A sample game played during an experiment on learning strategies. A human S played against a computer program that was also learning to play the game.

their behavior to perform well. The E² wished to test if this effect would appear if no one were "looking over the S's shoulder" and if it were absolutely clear to the S that his previous answers had not been recorded. This could be accomplished by

IT IS NOW WELL-ESTABLISHED THAT THE APPEARANCE OF A WORD TELLS CERTAIN THINGS ABOUT ITS MEANING. SPECIFICALLY, IN SOME LANGUAGES THERE ARE CERTAIN COMBINATIONS AND SPACINGS OF LETTERS THAT HAVE NEGATIVE CONNOTATIONS AND OTHERS THAT HAVE POSITIVE CONNOTATIONS. MODERN ENGLISH IS NOT ONE OF THESE LANGUAGES. HOWEVER, IF IT WERE, FOR EXAMPLE, THE "OO" IN GOOD MIGHT HAVE POSITIVE CONNOTATIONS THAT WOULD GENERALIZE TO OTHER WORDS, SUCH AS TOODE OR PROOD, SUCH THAT THEY WOULD ALSO HAVE POSITIVELY-TONED MEANINGS. OLD ENGLISH IS A LANGUAGE WHICH HAS THIS CHARACTERISTIC. THE ABILITY TO DISTINGUISH WHICH PATTERNS ARE POSITIVELY-TONED AND WHICH ARE NEGATIVELY-TONED CORRELATES HIGHLY WITH SENSITIVITY. IN THAT BOTH DEPEND ON RECOGNITION OF VERY SUBTLE CUES. OBVIOUSLY, PEOPLE DIFFER ON THIS TRAIT. SOME PEOPLE ARE VERY SENSITIVE TO SUBTLE CUES, WHEREAS OTHERS MAKE ONLY SUPERFICIAL JUDGMENTS. WE ARE INTERESTED IN SEEING HOW A COLLEGE POPULATION SCORES ON THIS TEST.

HAVE YOU STUDIED OLD ENGLISH? TYPE "YES" OR "NO."
>NO

TYPE (SEPARATED BY SPACES) YOUR LAST NAME, FIRST NAME, AND SEX (M OR F)
>DOE JOHN M

SHORTLY, YOU WILL BE PRESENTED WITH LISTS OF OLD ENGLISH WORDS. AFTER EACH WORD, YOU WILL SEE AN ARROW. TYPE IN "GOOD" IF YOU THINK THE WORD HAS POSITIVE CONNOTATIONS. TYPE "BAD" IF YOU THINK THE WORD HAS NEGATIVE CONNOTATIONS. TAKE AS MUCH TIME AS YOU NEED ON EACH WORD.

HASU
>GOOD

FEOND
>GOOD

WUNIAN
>BAD

MAECG
>GOOD

EACEN
>BAD

ASTIGAN
>BAD

YOUR SCORE IS 2 CORRECT OUT OF 6.

TWEGEN
>GOOD

BUGAN
>GOOD

LIEFAN
>GOOD

DAEL
>GOOD

SORG
>GOOD

PENDEN
>GOOD

YOUR SCORE IS 1 CORRECT OUT OF 6.

MEERAN
>GOOD

EOEL
>BAD

BLOD
>GOOD

AEROR
>GOOD

GIELLAN
>BAD

STRAEL
>GOOD

YOUR SCORE IS 2 CORRECT OUT OF 6.

*****ERROR: PL1167-I/O INTERRUPT RAISED. ERROR ON DISK ACCESS. ACCESS FAILS.

SORRY, DUE TO ELECTRONIC PROBLEM DATA ON LAST TEST LOST IN TRANSMISSION.
LAST TRIAL WILL BE REPEATED.

MEERAN
>GOOD

EOEL
>BAD

BLOD
>BAD

AEROR
>BAD

GIELLAN
>GOOD

STRAEL
>GOOD

THIS EXPERIMENT WAS NOT REALLY ABOUT SENSITIVITY. SOME OF YOU WERE GIVEN EITHER HIGH OR LOW SCORES ALL THE WAY THROUGH THE EXPERIMENT. WE DID NOT EXPECT YOU TO CHANGE TOO MANY ANSWERS WHEN GIVEN A CHANCE TO REPEAT YOUR LAST SIX ANSWERS, SINCE THEY WERE ABOUT WHAT YOU HAD EXPECTED. OTHERS OF YOU WERE GIVEN HIGH OR LOW SCORES UP UNTIL THE LAST SET WHEN YOUR SCORE WAS VERY DIFFERENT FROM THE PREVIOUS ONES. WHEN GIVEN A CHANCE TO CHANGE YOUR SCORES BY THE FAKE ERROR MESSAGE, WE EXPECTED YOU TO CHANGE THEM TO CORRESPOND WITH YOUR PREVIOUS PERFORMANCE, WHETHER HIGH OR LOW. WE COULD NOT TELL YOU THE TRUE PURPOSE AT THE OUTSET OR YOU WOULD NOT HAVE RESPONDED NATURALLY. YOU ARE REFERRED TO ARONSON AND CARLSMITH, IN THE JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY, VOLUME 65, 1962, STARTING ON P. 178. WERE YOU SUSPICIOUS? TYPE YES OR NO.
>YES

ALAS.

THE SUBJECT WAS IN THE LOW SELF-ESTEEM CONFIRMED EXPECTANCY CONDITION.
HIS NUMBER OF CHANGED ANSWERS WAS 3

==>PRINT DATAOUT
DOE JOHN M
LC 3
Y

END OF DATA SET
==>

Fig. 4. Part of a computer-controlled experimental session replicating the Aronson and Carlsmith (1962) experiment on disconfirmed expectancies.

controlling the experiment with a computer. Furthermore, as with all computer-controlled experiments, the possibility of E-induced bias would be greatly reduced.

A sample experimental session is shown in Fig. 4. After the S arrives, he is given instructions to the effect that "so many subjects are being run, I can't keep track of them. I don't have time to help you, but the computer running the experiment will provide all the help you need. Just go over to the teletype, press carriage return, and follow the instructions you get." The E then leaves the room. During an actual experiment, the S's responses do not print on the teletype so he cannot refer back to previous answers. The S is consistently told that he has done well or badly depending on the condition to which he has been assigned. A random-number generator is used to vary the exact scores that are fed back to him. Finally, after giving the S his score on what turns out to be the last set of questions, the computer hesitates and types an error message (see Fig. 4) saying that the last set of data has been lost due to an electronic problem. The S is given the same set of questions over again. Of course, there hasn't been any error and both sets of answers are saved. The S is then debriefed by the computer and by the E. The data can be analyzed immediately since it is already on the main computer.

This program was written in PL/I by the investigator in less than 2 days. Yet it was only the second program he had written. Such rapid construction of programs simply would not be possible on most experimental-control computers.

Example 3

The last example is an experiment on the effects of hunger on cognitive behavior. The researcher³ is faced with two problems. First, after a hunger manipulation he wishes to run a great many Ss. Since the hunger manipulation depends on elapsed time, it is important to run Ss in as short a time as possible. Second, the task on which all will be tested is selecting the "different" word from a triad. It was hypothesized that hungry humans would be more likely to attend to the attribute "edibility." Hence, the E wished *each* word in the triad to be "different" along *some* dimension, but only one word to be different along the edibility dimension. To construct such triads, the investigator proposed first scaling the stimulus words (animal names) on each dimension (edibility, domesticity, size, and hairiness) via a paired-comparison procedure and then constructing the triads from those scales. This could only be done in a reasonable amount of time if the experiment were controlled by a computer.

A program was designed that, with several terminals available, permits an E to run many Ss in one time period. Computer control produced the usual advantages of precise control of variables and immediate storage of data in the computer.

All the experimental programs described above were programmed by very inexperienced programmers and are running successfully on the Yale time-sharing system. Several other

programs have been written, including programs for verbal learning and one for a computer-controlled desensitization procedure.⁴ Of course, computer control does not guarantee that the experiment will turn out the way the investigator hopes; but it does minimize the chances of E bias, data-collection errors, and unintended fluctuations of independent variables.

CONCLUSION

For many experiments in the behavioral and biomedical sciences, there is no substitute for an experimental-control computer with true real-time capability. However, if an E can tolerate timing errors of several seconds, then he can reap the benefits of computer-controlled experimentation on a large general-purpose time-sharing system. Independent variables can be controlled precisely, the effect that the E's presence produces can be reduced, data collection can be completely automatic, and complicated calculations can be used to determine the course of the experiment. In addition, there are a few advantages in using the general-purpose computer instead of a small machine. Very sophisticated data analyses can be carried out on line and the data need not be transferred to a large computer. Programming effort is reduced since more convenient languages are available than on most small computers. But the biggest advantage is cost. The expense involved in constructing and maintaining a computer-controlled laboratory is large and cannot usually be justified for one or two experiments. Of course, a computer-controlled laboratory would be more economical than a large time-sharing system if a fair number of users were interested. Furthermore, many experiments cannot be run on a large time-sharing system, and others cannot be run well.

When only a few experiments are necessary, and a computer-controlled laboratory cannot be justified, investigators will do well to consider employing a large time-sharing system to control experiments.

REFERENCES

- ARONSON, E., & CARLSMITH, J. M. Performance expectancy as a determinant of actual performance. *Journal of Abnormal & Social Psychology*, 1962, 65, 178-182.
- BLOCK, H. D. Learning in some simple non-biological systems. *American Scientist*, 1965, 53, 59-79.
- GREEN, B. F., & HUESMANN, L. R. Computer time sharing. *Proceedings of the Personnel Research and Systems Advancement Symposium*, DDC:PRL-TR-67-13, December 1967, 57-66.
- JOHNSON, E. Concept formation experiments. *Psychometric Society meeting*, Princeton University, April 1969.
- UTTAL, W. R. *Real time computers*. New York: Harper & Row, 1968.

NOTES

1. C. Wood.
2. S. Taylor.
3. B. Cohen.
4. M. Storms.