10. SYMPOSIUM ON COMPUTER USE BY NONHUMAN ANIMALS: RESEARCH METHODS, TOOLS, AND SIMULATION

Chaired by David A. Washburn, *Georgia State University*

Simulations of schedule-controlled behavior

FERNANDO A. GONZALEZ
Morris Brown College, Atlanta, Georgia

A computer model that simulates the patterns of responding of infrahuman subjects under several schedules of reinforcement is described. The model is dynamic in that it continuously assesses the values of several interacting variables that are, in turn, affected by simulated environmental events that are scheduled by a procedure program. The data generated by the computer model, including cumulative records, closely conform to reported experimental data. The results indicate that computer simulations are a very useful tool for developing quantitative theories of operant behavior.

It has proven difficult to develop general quantitative models of operant behavior because so many interacting variables must be incorporated into the models. The complexity of the models expands rapidly as the number of included variables increases. Mathematical modelers, therefore, tend to develop simplified operant behavior models. But, because even the simplest of operant behaviors continuously interacts with many organismic, environmental, and historical variables, it is not easy to experimentally assess the adequacy of these streamlined models.

Fortunately, today's PCs, equipped with a few peripherals and loaded with a programming language, data-analysis package, and graphical software, allow research psychologists to create complex quantitative models and conduct computer-simulated experiments to explore the range of predictions of the model as the variables are manipulated. Any good PC can, in but a few minutes, run a simulation involving several concatenated nonlinear equations, and yield tens of thousands of data points representing the time-dependent state of many variables under differing conditions. Thus, computers make it possible for researchers to create complex models that both yield testable predictions and can be easily modified and expanded. Variables that cannot be controlled in an experiment can be manipulated in the model over their entire range. These exercises

can lead to obviously absurd predictions that invalidate the model, or to unexpected, testable predictions that permit the experimental assessment of the model's adequacy.

In this paper I present a computer model that simulates responding by infrahuman subjects under several schedules of reinforcement. The scope of this paper does not permit detailed discussions of the assumptions or equations that underlie this specific model. The purpose is to illustrate how computer models can simulate schedule-controlled responding, and how they can be used to help develop and test theories of operant conditioning.

THE PROGRAM

The simulation program consists of two interacting routines: a procedure routine and a subject routine. Values for procedure parameters and the operating characteristics (OCs) of the subject are provided by the researcher at the start of the simulation. As the simulation proceeds, some of the OCs are affected by exposure to the procedure simulation. At the end of each session, the OCs are automatically dumped into data files and then fed into the subject routine at the beginning of the next session. The program continuously records information on the subject's behavior, the OCs, and procedure events. These data are then used to generate cumulative records and summary statistics. The model predicts the details of transitions and the asymptotic characteristics of responding as the variables are manipulated. For the purposes of this paper, only schedules and parameter values of schedule components were manipulated. Initial values of the OCs were the same in all the simulations reported here; they are presented in the Appendix.

This research was supported by a grant from the National Institutes of Mental Health to Morehouse College (R24 MH 47188-02) and from the U.S. Army Research Office to Morris Brown College (DAAL03-92-G-0382). Reprints and copies of the program may be addressed to F. A. Gonzalez, Center of Excellence for Research on Training, Morris Brown College, 643 MLK Jr. Dr., NW, Atlanta, GA 30314.

Figure 1. Process chart for procedure program. Ra, Rb, Rab, and Rba denote reinforcers for schedules VTa, VTb, VTab, and VTba, respectively. Ta, Tb, Tab, and Tba denote time intervals during which the corresponding schedules are in effect. Tcod denotes the time in a changeover delay (COD). Dashed lines indicate process paths specific to each schedule. Procedure parameters for CONC VTVT: VTa = VTba, VTb = VTab; parameters for CONC VrTVT: VTa = very large, VTba = VrT parameter value, VTb = VTab; parameters for VI or VR: VTa = very large, VTba = VI or VR parameter value, VTab = very large, VTb = small. (A slight adjustment in the formula for assigning reinforcers is also required for the VR simulation.)

Procedure Routine

A flow chart of the procedure routine is presented in Figure 1. The basic procedure routine simulates a concurrent variable time, variable time (CONC VTVT) schedule (Brownstein & Pliskoff, 1968). This schedule offers the subject two behavior alternatives, designated in the chart as Behaviors A and B. Access to A or B at the start of each session was selected at random. In CONC VTVT schedules, a specific short-duration response is not required for presentation of the reinforcer. Instead, when the subject selects an alternative, say Behavior A, then RA reinforcers are presented at a constant rate VT_A as long as a changeover (CO) to the other alternative does not occur, and one R_{AB} reinforcer is assigned at a constant-rate VT_{AB} for delivery after a fixed changeover delay (COD) following the changeover to B. Therefore, the probability that an R_{AB} reinforcer will be assigned during an episode of Behavior A increases as a decelerating exponential function of the duration of the episode of A. If assigned, the R_{AB} reinforcer is presented immediately after the COD fol-

lowing the CO from A to B. The VT_B schedule for reinforcement during B is then initiated, as well as the VT_{BA} schedule for presentation of an R_{BA} reinforcer following the CO back to A.

Usually, in CONC VTVT schedules, VT_A = VT_{BA} and VT_B = VT_{AB}. Most of the published experimental results have been obtained with such schedules. This is not necessary, however, and in the present model the four programmed rates can be manipulated independently. A few modifications to the procedure program and some changes in procedure parameter values are all that is needed to convert the CONC VTVT simulation to the other schedules considered in this paper—concurrent time-based variable ratio, variable time (CONC VrTVT) (Heyman & Herrnstein, 1986); variable interval (VI); and variable ratio (VR). The dashed lines in the flow chart denote the required alterations to the CONC VTVT program.

The parametric changes in the procedure variables that are needed to convert the CONC VTVT to one of the other schedules are indicated in the caption to Figure 1. For ex-

An assumption of the behavior model is that the process of selecting between alternatives takes time. It is also assumed that the distribution of interdecision intervals (IDIs) is exponential; thus, mean IDI = $1/DR$, where DR is the decision rate. In the following examples, DR is set to 1.5 decisions per second. This sets the ceiling for mean CO rates. Whether or not a CO occurs depends on the output of the preference equation. Equation A3 in the Appendix yields the preference (PRE) for changing over from one behavior to the other.

Notice that the range of Equation A3 is -1 to 1 . The decision rule is that if PRE is positive, then a CO takes place with probability equal to PRE, but if PRE is negative, then the probability of CO is zero. If a decision does not take place, or if a CO fails to occur, then the behavior routine advances the appropriate timer, checks for reinforcer presentation, reassesses the expectations, and re-initiates the decision process. As time elapses and the decision rule is not applied, the probability that it will be applied during the next IDI increases exponentially, approaching the value of one. Similarly, with the CONC VTVT or VI procedures, as time elapses and a CO or a response does not occur, the expected probability of a reinforcer following the next CO or response increases rapidly, assuring that PRE will become positive and that the CO or response will eventually occur.

With the CONC VrTVT and VR schedules, however, the probability of presentation of scheduled reinforcers is independent of the interchangeover and interresponse times. The flow chart in Figure 1 shows that in the VrT component of the CONC VrTVT as well as in the VR schedule, the time-based reinforcement loop, indicated by the dashed line labeled 1, is bypassed (by redirection through dashed line 3). In these schedules, the value of the utilities changes only when the value of the expectations of reinforcement rates changes.

Results

The simulation program is run from a batch file that permits specification of multiple sessions and conditions. All simulated sessions last 1 h (computer time). With a 486 PC at 50 MHz, it takes approximately 25 sec to run one session, transcribe the data to the cumulative record data file, and complete preliminary data analyses. A statistics and graphics package was used for cross-session data analyses and plotting.

Only a limited portion of the results can be presented in this article due to length restrictions. Data obtained with the VI and VR simulations were selected for display because readers are more likely to be familiar with the characteristics of responding maintained with these schedules.

Figure 3 shows that the simulation of the VR schedule yielded higher rates of responding than the VI simulation at any given rate of obtained reinforcers. The results also indicated that with VR and VI parameter values that maintained comparably low response rates, session-to-session variability in responding was much greater with the VR than with the VI simulation.

Figure 3. Response rate (y-axis) as a function of reinforcement rate (x-axis) for VR 15, 30, 60, and 120, and VI 15, 30, 60, 120, and 240 sec. Each condition remained in effect for 20 1-h simulation sessions. Each data point represents the average response rate from a session. Curves are rectangular hyperbolae fitted by iteration to no more than 5% error (chi-square).

The cumulative records in Figure 4 show that responding was similarly consistent during VR and VI simulation sessions in which response and reinforcement rates were high (VI 20 sec, VR 20). However, with the lower densities of reinforcement (VI 240 sec, VR 60), response rates were more variable under the VR than they were under the VI. Furthermore, with the lower reinforcement rates, VR responding became irregular or "strained" (Ferster & Skinner, 1957)—the cumulative records indicated abrupt changes in rate and frequent pauses from responding, followed by short periods of high response rate. At yet lower reinforcement rates (not shown), the pauses became more frequent and prolonged, and eventually responding ceased.

Distinctions between VI and VR transitions are clearly seen in the cumulative records for VI Sessions 61 and 81 and VR Sessions 61 and 90, which correspond to the first extinction sessions and to the recovery sessions. Note that VI recovery occurred almost immediately, whereas VR responding did not recover until the 10th session after the VR 20 had been reinstituted. This is as expected, given the response-reinforcement contingencies specified by the schedules.

DISCUSSION

The simulations presented here are based on a model of operant behavior that incorporates procedure variables (schedules, schedule parameter values, COD), organismic

SESSION No. 20 VI= 20 s

SESSION No. 20 VR= 20

SESSION No. 40 VI= 240 s

SESSION No. 40 VR= 60

SESSION No. 61 VI= 1000000 s

SESSION No. 61 VR= 1000000

SESSION No. 81 VI= 20 s

SESSION No. 90 VR= 20

Figure 4. Representative cumulative records of responding under VI (left column) and VR (right column) schedule simulations.

constants (EF, CF, SC, MOT, EFF, d), and historical variables (expectations of reinforcement rates, confidence intervals of the expectations). The historical variables and the constants are used as terms in equations to predict behavior. They are analytic constructs, without implied physiological or psychological reality. The names given to the terms were selected to communicate a sense of the role played by the constants or variables in the model.

The behavior model can be classified as a momentary maximization model (Shimp, 1966). It is *dynamic* (Staddon, 1988) in that it predicts behavior during consecutive, brief intervals; it is also, to some extent, *general*, since it accommodates several important experimental procedures. The output of the model—that is, the simulated behavior—is driven by an environmental variable: the presentation of reinforcers. The expectations of reinforcement rates and the confidence associated with the expectations are states of the “subject” that reflect its history of interaction with reinforcer presentations. The constants represent characteristics of the subject that limit the magnitude of interactions between variables. Most of the observed variability reflects the changing expectations of reinforcement rates; the remaining variability results from the exponential distribution of IDIs. The formula for determining IDI durations was selected principally on the basis of computational ease. Similarly, because motivation, effort, and reinforcement delay were not crucial to the simulations, these factors were incorporated simplistically into the model. The formulae of their inclusion are not intended to suggest a quantitative theory of how they influence operant behavior.

The hypothesis-testing component of this model is more typical of cognitive models of decision making than of behavioral models of choice. The first few versions of the model were based on Herrnstein's notion of melioration (Herrnstein, 1982) and did not include the hypothesis-testing component. This element was added on a whim to “see what would happen if.” That modified version did not, at first, include the shifting CIs. The hysteresis functions were added when it was discovered that the shifting CIs were unstable and “crashed” to extreme values. The decision rule was also modified several times as the model evolved. The rapid evolution of this model denotes an advantage of computer modeling: ease of modification. Student assistants working on this project have experimented freely with many features of the model. This has proven to be an excellent means to introduce them to, and interest them in, computer modeling.

Predictions of the behavior model conform well with reported details of experimental data. The model also yields unexpected (and unambiguous) results with variants of the CONC schedules that have not been used in experiments. On the other hand, the model has two limitations that severely restrict its generality: It only applies to procedures involving exactly two alternative behaviors, and it does not include a subroutine for temporal discrimination. Within these limitations, it seems to work quite well. The aim of this paper, however, has not been to defend this specific model, but to illustrate how computer

modeling can be an effective tool to help develop quantitative theories of behavior.

REFERENCES

- BROWNSTEIN, A. J., & PLISKOFF, S. S. (1968). Some effects of relative reinforcement rate and changeover delay in response-independent concurrent schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, 11, 683-688.
- FERSTER, C. B., & SKINNER, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.
- HERRNSTEIN, R. J. (1982). Melioration as behavioral dynamism: Parallels and intersections. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 433-458). Cambridge, MA: Ballinger.
- HEYMAN, G. M., & HERRNSTEIN, R. J. (1986). More on concurrent interval-ratio schedules: A replication and review. *Journal of the Experimental Analysis of Behavior*, 46, 331-351.
- SHIMP, P. C. (1966). Probabilistically reinforced choice behavior in pigeons. *Journal of the Experimental Analysis of Behavior*, 9, 443-455.
- STADDON, J. E. R. (1988). Quasi-dynamic choice models: Melioration and ratio invariance. *Journal of the Experimental Analysis of Behavior*, 49, 303-320.

APPENDIX

The following difference equation is used in the behavior program to adjust the values of the expectations of reinforcement rate when the observed probabilities of reinforcement fall outside the confidence intervals of the expected probabilities:

$$E_{x,n+1} = E_{x,n} \pm \left[EF \times E_{x,n} \times (1 - E_{x,n}) \times e^{-\frac{SC \times CIE_{x,n}}{1 - CIE_{x,n}}} \right]. \quad (A1)$$

$E_{x,n}$ is the expectation for reinforcer R_x during the n th 0.1-sec interval of the session; the expectation change factor ($0 \leq EF \leq 1$) is a constant that establishes the maximum rate of change of the expectations; the exponential term relates the rate of change to the confidence interval for the expectation (CIE_x), and the sensitivity coefficient ($0 < SC < 1$) determines the value of the exponential term as CIE_x varies. Note that as CIE_x approaches one, the exponential term approaches zero, and therefore the magnitude of changes in expectation also approaches zero.

Changes in the confidence intervals of the expectations of reinforcement rates when the expectations are not confirmed are determined by the following difference equation:

$$CIE_{x,n+1} = CIE_{x,n} \pm \left[CF \times CIE_{x,n} \times (1 - CIE_{x,n}) \times e^{-\frac{(PK - CIE_{x,n})^2}{TL - (1 - CIE_{x,n})}} \right], \quad (A2)$$

where the confidence change factor ($0 \leq CF \leq 1$) is a constant that determines the maximum rate of change. The constants PK (peak) and TL (tails) determine the peak and the spread of the hysteresis curves. PK and TL are assigned slightly different values for increases and decreases in confidence. The exponential term keeps the CIE_x from approximating—on the high side—the value of 1, which would make it impossible to reject the null hypothesis that the observed rate of reinforcement is

not different from the expected rate, and also keeps CIE_x from approximating 0.75 on the low side, which would make it very unlikely to fail to reject the null hypothesis.

Decision Rule

$$PRE_{COxy} = \frac{U_{xy} - U_x}{U_{xy} + U_x} \quad (A3)$$

The terms in Equation A3 represent the “utility” to the subject of continuing to engage in an alternative (U_x), and of changing over to the other (U_{xy}). The decision rule for changeovers is as follows:

$$P(CO) = \begin{cases} PRE & \text{if } PRE > 0 \\ 0 & \text{if } PRE \leq 0 \end{cases}$$

Equation A4 denotes that, at any given moment, the value of the utility of a CO from x to y is a function of (1) the estimated frequency of reinforcement, given E_{xy} and E_y , during the mean IDI immediately following the CO; (2) the estimated frequency of reinforcement, given E_x , during the next mean IDI if there is no CO; (3) the motivations ($MOT \geq 0$) associated with the reinforcers; (4) the effort ($EF \geq 0$) associated with each behavior and CO; and (5) the delay of reinforcement effect associated with the COD ($d \geq 0$).

$$U_{xy} = \left[(1 - e^{-E_{xy} \times T_x}) \times \frac{MOT_{xy}}{1 + EFF_{xy}} + IDI \times E_y \times \frac{MOT_y}{1 - EFF_y} \right] \times e^{-d \times COD} \quad (A4)$$

Equation A5 describes the utility associated with continuing the ongoing behavior—that is, with not changing over.

$$U_x = IDI \times E_x \times \frac{MOT_x}{1 - EF_x} \quad (A5)$$

Initial values of the expected rates of reinforcement (E) were made equal to the corresponding scheduled rates of reinforcement, and the initial confidence intervals (CI) were set to 0.95. In the CONC schedules simulations, $EFF_A = EFF_B = 0.05$ and $EFF_{AB} = EFF_{BA} = 0.075$, $MOT_A = MOT_{BA} = MOT_B = MOT_{AB} = 1$, and $COD_{AB} = COD_{BA} = 2$ sec. In the VR and VI simulations, $EFF_{BA} = 0.075$, $EFF_A = 1$, $EFF_B = 0$, and $EFF_{AB} = 0$; $MOT_{BA} = 1$, $MOT_A = 0$, $MOT_B = 0.05$, $MOT_{AB} = 0$, $COD_{AB} = 0.01$ sec and $COD_{BA} = 0.05$ sec. For all the schedules, $EF = 0.25$, $CF = 0.1$, $SC = 0.05$, $d = 0.05$; $PK = 0.925$, and $TL = 0.05$ for contraction of CI , and $PK = 0.85$ and $TL = 0.1$ for expansion.