SESSION III SIMULATIONS OF PSYCHOLOGICAL MODELS

Michael Yost, *Chair*

A computer simulation/mathematical model of learning: Extension of DMOD from appetitive to aversive situations

HELEN B. DALY and JOHN T. DALY

State University of New York College at Oswego, Oswego, New York

The computer simulation/mathematical model called DMOD, which can simulate over 35 different phenomena in appetitive discrete-trial and simple free-operant situations, has been extended to include aversive discrete-trial situations. Learning (V) is calculated using a three-parameter equation $\Delta V = \alpha\beta(\lambda - \bar{V})$ (see Daly & Daly, 1982; Rescorla & Wagner, 1972). The equation is applied to three possible goal events in the appetitive (e.g., food) case and to three in the aversive (e.g., shock) case. The original goal event can be present, absent, or reintroduced; in the appetitive situation, these events condition approach (V_{ap}), avoidance (V_{av}), and courage (V_{cc}), respectively. In the aversive situation, the events condition avoidance (V_{av}^*), approach (V_{ap}^*), and cowardice (V_{cc}^*), respectively. The model was developed in simple learning situations and subsequently was applied to complex situations. It can account for such diverse phenomena as contrast effects after reward shifts, greater persistence following partial than following continuous reinforcement, and a preference for predictable appetitive and predictable aversive events. Application of the aversive version of the model to "reward" shifts is described.

Our goal is to develop a computer simulation/mathematical model of learning that is as simple as possible with as much breadth as possible. The model, called DMOD (Daly MODification of the Rescorla-Wagner Model), is simple because learning is calculated with one simple equation using three parameters. It is applied to diverse situations by assuming that there are a number of different goal events possible, each of which conditions either approach or avoidance of the goal. It was originally developed to account for appetitive, discrete-trial experiments, and can currently account for behavior in over 30 different paradigms (see Daly & Daly, 1982). DMOD was then extended to simulate the effect of simple schedules of reinforcement in free-operant experiments (Daly & Daly, 1984b). Our purpose is to outline the extension of DMOD to the aversive case and to show its application in a complicated experimental situation. To understand the rationale behind the extension, however, it is neces-

sary to review development of the model in the appetitive case and the rules we follow for development of DMOD.

SELECTION OF PHENOMENA TO BE SIMULATED

We believe that the initial goal of a new model is to be able to account for well-established and replicable phenomena. We feel that it is dangerous to develop a model around recently discovered phenomena, because the boundary conditions under which they can be obtained and the variables influencing them are unknown. The purpose of a theory, however, is not only to integrate existing replicable data, but also to correctly predict new results. The primary phenomenon we use to test predictions of the appetitive version of DMOD is acquisition of a preference for predictable reward. The model made some interesting predictions concerning when a preference for the unpredictable reward situation should occur. These predictions were subsequently supported (Daly, 1985; Daly & Daly, 1982); other theories have great difficulty accounting for these results (Daly, 1985, 1986). We are just beginning to use the extensions of the model

Preparation of this article was supported in part by National Institute of Mental Health Grant MH33497 to Helen Daly. Requests for reprints should be sent to Helen Daly, Department of Psychology, or John Daly, Department of Mathematics, State University of New York College at Oswego, Oswego, NY 13126.

to predict new results in free-operant and aversive situations.

COMPUTER PROGRAM

In 1984, Daly and Daly published the computer program and instruction manual for DMOD applied to appetitive, discrete-trial experiments. The computer program, written in BASIC, is easy to use: one simply has to enter values for the parameters and define which stimuli are present on which trials. Trial-by-trial changes in learning and values at asymptote are printed. The instruction manual has several sample simulations. Additional descriptions of the computer program can be found in Steinhauer (1986).

We have since written a much faster, menu-based program in the programming language C. It is run on Digital Equipment Corporation (VAX and PDP-11) and AT&T (7300 UNIX PC) machines, but it will run on most microcomputers with a C compiler. The program can be used to simulate appetitive discrete-trial and simple free-operant situations and discrete-trial aversive situations. (The program will be made available in August, 1987.)

PRECISE SPECIFICATION AND QUANTITATIVE PREDICTIONS

There is substantial empirical evidence to conclude that simple learning is an incremental process. We use the Rescorla-Wagner (1972) version of the linear-operator model: $\Delta V = \alpha\beta(\lambda - \bar{V})$, where α (α) is the intensity of the stimulus, β (β) is the rate of learning, λ (λ) is the size of the goal event (US or reward magnitude), and $\bar{V}$ is the total amount learned to all stimuli present. This equation has been highly successful in accounting for and predicting many phenomena (Rescorla & Wagner, 1972) and is currently being used to account for behavior in complex cognitive tasks (e.g., Gluck & Bower, 1986) and for learning at the neural level (e.g., Gluck, Donegan, & Bower, 1986).

The parameter values, α , β , and λ , are fixed by the experimental procedures: λ is determined by the size of the goal event, such as food; β by the vigor of the goal response, which is determined by the hunger drive level if food is the reinforcer; and α by the intensity of the stimulus. Much work needs to be done to scale each of the parameters, and, until this work is done, the model is restricted to making ordinal predictions; it predicts the ordering of the groups rather than precise differences between groups. The same parameter values are used to simulate different experiments that use similar values; λ is 1, .1, and 0 for large, small, and no reward; β is .15 if rats are reduced to 80% of their ad-lib weight; and α is 1 for all stimuli, which effectively eliminates this parameter. Only when stimulus intensity is varied is a lower α value used for lower intensity stimuli. Relative ordering of groups is rarely influenced by changing the

parameter values; only the magnitude of group differences is influenced.

DEVELOPMENT OF DMOD: APPETITIVE CASE

Whenever a positive reward is given, approach strength (Vap) is calculated using the ΔV formula listed earlier:

$$\Delta V_{ap} = \alpha\beta(\lambda - \bar{V}_{ap}).$$

This formula has been highly successful in accounting for phenomena, such as blocking, overshadowing, and super-normal conditioning (see Rescorla & Wagner, 1972). It is, however, unable to account for behavior in many situations, especially when reward values are changed. For example, it cannot account for the "depression" effect: subjects who were shifted from a large to a small reward approach the small reward more slowly than those who experienced only small rewards. The formula also cannot account for the partial reinforcement extinction effect: subjects rewarded on every trial during acquisition extinguish faster than subjects given extinction training following partial reinforcement.

The Rescorla-Wagner model was modified because of the phenomena it could not simulate. We assumed that two additional important goal events can occur in appetitive conditioning situations. For example, once one has a fairly large Vap value and a response is not followed by a positive reward (a most unpleasant experience), one learns to avoid the goal. The strength of the avoidance response is calculated using the same formula, except that the λ value is different. The λ value is the difference between the size of the reward obtained (0 or nothing on an extinction trial) and what was expected, which is indexed by Vap. Therefore, avoidance strength (Vav) is calculated using the formula:

$$\Delta V_{av} = \alpha\beta[2(0 - \bar{V}_{ap}) - \bar{V}_{av}].$$

Vav is negative, and the more negative the number, the greater the avoidance strength. Other researchers have shown that the loss of something (e.g., \$100) influences a subject more than its gain (e.g., Kahneman & Tversky, 1982). Therefore, the λ value is multiplied by a value greater than 1. We are using the value 2 until additional research suggests otherwise. Additional nonrewarded trials increase Vav and decrease Vap. Vap is decreased because the λ value to calculate Vap is zero on a no-food trial [$\Delta V_{ap} = \alpha\beta(0 - \bar{V}_{ap})$]. Additional rewarded trials decrease Vav (becomes less negative) and increase Vap. If the reward is reintroduced after Vav is conditioned, counterconditioning occurs (Vcc). It is calculated using the formula:

$$\Delta V_{cc} = \alpha\beta[(0 - \bar{V}_{av}) - \bar{V}_{cc}].$$

This is the same basic equation, except the λ value is the discrepancy between the aversiveness experienced (0 on

a positive reward trial) and the amount of aversiveness expected, which is indexed by V_{av} . One can view V_{cc} as courage conditioned to approach an aversive goal. Behavior is assumed to be the sum of the V values conditioned: $V_t = \bar{V}_{ap} + \bar{V}_{av} + \bar{V}_{cc}$. The addition of V_{av} allows the model to account for the "depression" effect, and V_{cc} allows the model to account for the partial reinforcement extinction effect (see Daly & Daly, 1982).

With additional assumptions about the effects of secondary reinforcement, delay of reinforcement, spatial generalization of aversive gradients, threshold for aversiveness, and effects on β values of certain drugs, the model is able to account for over 30 different phenomena (see Daly & Daly, 1982). Mathematical derivations of the asymptotic V values on a partial reinforcement schedule show that in complex situations the model is no longer linear (see Daly & Daly, 1982, Appendix). Extension of the model to the free-operant situation, where the V_t value is used to predict the time to the next response (interresponse time or IRT), allows it to account for acquisition and asymptotic response patterns under FI, VI, FR, and VR schedules of reinforcement (Daly & Daly, 1984b). Work is currently under way to extend the model to more complex schedules of reinforcement.

DEVELOPMENT OF DMOD: AVERSIVE CASE

Assume that a subject receives a positive reward on every trial, but is also given an electric shock. Receiving a shock is an aversive event, and the subject learns to avoid the goal: V_{av}^* is conditioned. (We use the * to differentiate an avoidance gradient based on shock from an avoidance gradient based on the absence of an appetitive event.) The formula used to calculate V_{av}^* is:

$$\Delta V_{av}^* = \alpha\beta[\lambda^* - \bar{V}_{av}^*],$$

where the λ value is determined by the intensity of the shock. This equation is identical to the Rescorla-Wagner model. It also cannot account for a number of phenomena. Therefore, we assume that there are additional important goal events. If V_{av}^* is conditioned and shock does not occur, this is a positive event, and V_{ap}^* is conditioned:

$$\Delta V_{ap}^* = \alpha\beta[(0 - \bar{V}_{av}^*) - \bar{V}_{ap}^*].$$

The λ value is the difference between the aversiveness experienced (0 or none when there is no shock) and the amount of aversiveness expected, which is indexed by V_{av}^* . Once V_{ap}^* has been conditioned and the subject is suddenly given a shock, this is another aversive event, and V_{cc}^* is conditioned:

$$\Delta V_{cc}^* = \alpha\beta[2(0 - \bar{V}_{ap}^*) - \bar{V}_{cc}^*].$$

The λ value is the discrepancy between the "relief" experienced (0 or none on a shock trial) and the amount expected, which is indexed by V_{ap}^* . It, too, is multiplied by 2, to be consistent with the assumption that aversive events are more important than positive ones. One can

Table 1
V Values Conditioned for Different Types of Goal Events

	Goal Event		
	Present	Absent	Reintroduced
Food	V_{ap} +	V_{av} -	V_{cc} +
Shock	V_{av}^* -	V_{ap}^* +	V_{cc}^* -

+ Subject approaches goal event. - Subject avoids goal event.

view V_{cc}^* as "cowardice" to approach. Behavior in aversive conditioning situations is determined by the sum of the V values: $V_t^* = \bar{V}_{av}^* + \bar{V}_{ap}^* + \bar{V}_{cc}^*$.

In summary, there are three important goal events possible in both appetitive and aversive conditioning situations which condition different V gradients. These are summarized in Table 1. The V_{ap} and V_{av}^* values are identical to the Rescorla-Wagner model and are the only V values conditioned in simple learning experiments where the same λ value is present on every trial. Only when the λ value is decreased do V_{av} or V_{ap}^* become conditioned, and only when the λ value is then increased after it was decreased do V_{cc} or V_{cc}^* become conditioned. Therefore, not all V values are conditioned in all experiments.

APPLICATION OF DMOD-AVERSIVE

We have successfully simulated 15 different phenomena in aversive conditioning situations. One relatively complicated experiment is presented here to demonstrate how the model is applied to escape and avoidance conditioning experiments.

We assume that *all learning involves approach to a goal*, and V values determine the strength of behavior toward that goal. In appetitive situations, the V_t value determines speed of approaching the goal. In punishment situations, V_t^* and V_t values are added. In simple shock escape/avoidance experiments, subjects must also learn to approach a goal, the goal now being a place with less aversiveness. For example, if subjects (typically rats) must learn to jump from the shockbox to a safebox to either avoid or escape a shock, the cues associated with shock in the shockbox have V_{av}^* conditioned to them. The reduction in aversiveness experienced when the subject reaches the safebox is assumed to be the λ value to condition V_{ap}^* . On the first trial the subject probably makes numerous types of responses, probably in some hierarchy of species-specific defense reactions and behavior learned in similar situations. As soon as a response is successful in reaching a less aversive place, approach (V_{ap}^*) is conditioned. On subsequent trials, the V_{ap}^* value determines the speed of approach to the safebox.

McAllister, McAllister, Brooks, and Goldman (1972) showed that speed of approaching a safebox is determined by the reduction in aversiveness and that speed is greatly depressed if subjects are shifted from a large to a small reduction in aversiveness. They gave rats 35 light-shock

pairings in a white box with a grid floor. Subjects were not allowed to escape from the box. On the next day, the animals were given no further shock trials, but were placed into the white box with the grid floor, the light was turned on as a door was opened, and the subjects could jump into a safebox. This box was black with a solid wood floor, and it was unlit. This was the "large reward" condition. The safebox for another group was white with a grid floor, and a light was on. This was the "small reward" condition. Both groups showed an increase in speed of jumping to the safebox with continued training, but the large reward group jumped more quickly (mean reciprocals of latency were about .48 vs. .28 on the last four preshift trials for the large and small reward groups, respectively; taken from McAllister et al., 1972, Figure 3, p. 498). To simulate this result, the following equation was used to calculate V_{ap}^* in the safebox:

$$\Delta V_{ap}^* = \alpha\beta(\text{reduction in } V_{av}^*) - \bar{V}_{ap}^*.$$

The reduction in V_{av}^* is the λ value, and is the difference in amount of aversiveness aroused in the safebox minus the amount aroused in the shockbox. In the large reward group, there is no aversiveness aroused in the safebox, so λ is $0(\text{safebox}) - V_{av}^*(\text{shockbox})$. In the small reward condition, there is V_{av}^* in the safebox because of the similarity between the two boxes. If one assumes that one half of the V_{av}^* is present in the safebox, the

λ value is only half that of the large reward group, since only half of the aversiveness is reduced. The smaller λ level results in a smaller V_{ap}^* value conditioned in the safebox, and therefore the group was expected to jump more slowly to the safebox. McAllister et al. (1972) then shifted half the subjects in each group to the other reward condition, and subjects shifted from the large to the small reward showed a quick drop in speed below the level of the control group, which had always experienced the small reward (.33, .28, .20, .20, .11, .11, .12, and .09 for the eight shift trials, values well below the small reward control group, which showed a slight increase to about .40; taken from McAllister et al., 1972, Figure 3). This is the precise condition where V_{cc}^* is conditioned (see formula previously given). V_{cc}^* is a negative number and decreases the speed in approach to the safebox. Therefore, DMOD can simulate this result. Another group was shifted from the small to the large reward and showed an increase in speed, but did not jump more quickly than the group always given the large reward. In fact, the means were lower on all but two trials. Simulations of DMOD show that V_{ap}^* grows following the shift, but there is no additional V term conditioned. Therefore, the shifted group should not jump faster than the unshifted groups. Figure 1 shows the DMOD simulation results. They are very similar to those obtained by McAllister et al. (1972, see Figure 3). Parameter values influence the size of the group differences, but do not change the pattern of results.

Figure 1. DMOD simulation results of McAllister et al.'s (1972, Experiment 3) reward shift experiment of jump speed to escape cues previously paired with shock, for the last three trials of large (L) or small (S) rewards (reduction in aversiveness following the jump), and eight shift trials, where two groups were shifted to the opposite reward condition (e.g., L-S: large reward Trials 1-11, small reward Trials 12-20). [Parameter values: $\lambda(\text{shock}) = 1.0$, $\beta_1 = .15$ (used when V values increment), $\beta_D = .05$ (used when V values decrement), $\alpha = 1.0$.]

CONCLUSION

We believe that the extension of DMOD from the appetitive to the aversive case is the logical one, given the assumption that goal events are either present, absent, or reintroduced, and that approach and avoidance strengths are conditioned. Applications of the extension to existing data have been successful, but complete simulations of all experimental manipulations have not been done. Work on the model will continue before releasing the computer program (hopefully by August, 1987). We feel it important, however, to present the extension to obtain feedback from researchers at this stage of model development.

REFERENCES

- DALY, H. B., & DALY, J. T. (1982). A mathematical model of reward and aversive nonreward: Its application in over 30 appetitive learning situations. *Journal of Experimental Psychology: General*, *111*, 441-480.
- DALY, H. B., & DALY, J. T. (1984a). DMOD—A mathematical model of reward and aversive nonreward in appetitive learning situations: Program and instruction manual. *Behavior Research Methods, Instrumentation, & Computers*, *16*, 38-52.
- DALY, H. B., & DALY, J. T. (1984b, November). *Extension of DMOD from discrete-trial to free-operant: Progress Report #1*. Paper presented at the Psychonomic Society Meeting, San Antonio, TX.
- DALY, H. B. (1985). Observing response acquisition: Preference for unpredictable appetitive rewards obtained under conditions predicted by DMOD. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 294-316.
- DALY, H. B., (1986). Dinsmoor's selective observing hypothesis probably cannot account for preference for unpredictable rewards: DMOD can. *The Behavioral & Brain Sciences*, *9*, 365-370.
- GLUCK, M. A., & BOWER, G. H. (1986, November). *Category learning, judgment, and the Rescorla-Wagner model (aka the delta-rule)*. Paper presented at the Psychonomic Society Meeting, New Orleans, LA.
- GLUCK, M. A., DONEGAN, N. H., & BOWER, G. H. (1986, November). *Computational models of the neural bases of associative learning*. Paper presented at the Psychonomic Society Meeting, New Orleans, LA.
- KAHNEMAN, D., & TVERSKY, A. (1982). The psychology of preferences. *Scientific American*, *246*, 160-173.
- MCALLISTER, D. E., MCALLISTER, W. R., BROOKS, C. I., & GOLDMAN, J. A. (1972). Magnitude and shift of reward in instrumental aversive learning in rats. *Journal of Comparative & Physiological Psychology*, *80*, 490-501.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.
- STEINHAUER, G. D. (1986). *Artificial behavior*. New Jersey: Prentice-Hall.