DYNAMIC PROGRAMMING AND PSYCHOLOGY

GRAHAM M. TAYLOR

Department of Psychology and Sociology,
University of Canterbury

The use of functional equations to describe behaviour is demonstrated using simple examples. A perceptual maze problem is analyzed using a model based on some algorithms of dynamic programming. The model is applicable to work on functional brain damage in alcoholics.

The key notions of dynamic programming are concerned with the control of behaviour. The approach is exemplified by a classical problem in behaviour analysis brought to my attention by Professor Gregson. In this example, we have a bug in a box. The environment of the box is static. Mathematically, if we call the parameters of the environment π : i, where $i = 1, 2 \ldots n$, then $d\pi : i/dt = O$, for all i.

We now introduce a bug into this sterile environment. To keep the analysis simple I will neglect any problems a bug may encounter from sensory deprivation. Bug moves at a constant speed, and his rate of turning in degrees per unit distance travelled forward is a linear function of the illumination level at the point where he is. Translating this into algebra, we have:

$$d\,\Phi/dl\ 1 = 1 : 1' \qquad (1)$$

where ϕ is the angle of turn, l is the length of path segments, and μ is the illumination level at 1′. We make l small enough such that μ is constant over 1′ no matter what the direction of travel. We may note from equation (l) that if bug is in total darkness he will go in a straight line; if the box is brightly lit he will rotate in a circle of zero radius.

We now put bug in a box of finite width and infinite length and let the illumination gradient across the box from wall to wall increase uniformly from no light at the left-hand side to very bright at the right-hand side. The illumination gradient does not change along the length of the box.

If we initially place bug at x units of length from the left-hand wall, we can now replace equation (1) by the following:

$$d\,\phi\ /\ d\,l = x \qquad (2)$$

If bug is put in front of a psychologist who knows nothing of differential equations then we have a high probability of getting from that psychologist a catalogue of behaviours labelled wall-hugging, circling, stopping on the spot, and so forth. Behaviour will be controllable in the sense used by the operant psychologist, and he may say 'alter the light level and you have altered reinforcement'. But the whole point of this example is this — bug hasn't got a repertoire of responses,

he has only one. We can predict what bug is going to do from now to eternity given his boundary conditions and his starting point.

This example shows that in some situations it makes more sense to work with models which account for behaviour, and to forego traditional ways of 'explaining' the data.

We may take this approach a step further, which is what Toda (1962) did in his design of a fungus eater. Those of you who know this work will remember that Toda set out to design a fungus eater. This fungus eater was placed on a planet which had high levels of uranium, in the form of pure nuggets strewn over the surface of the planet in a systematically varying fashion. Also on this planet grew a primitive fungus, whose distribution over the planet's surface also varied systematically. The task of the fungus eater was to collect uranium for the designer who sat back on earth, but in order to collect the ore the fungus eater must search for and consume fungus, its only source of energy. We may build into such a situation a number of well-defined constraining conditions on the behaviour of the fungus eater. Costs on his survival will play a part in predicting what fungus eater does in a given situation. What Toda is suggesting is that we begin with an environment and attempt to design a subject with the minimal optimal qualities to function effectively in this environment. This is close to the real world situation where efficiency is not measured by the speed with which a man presses a buzzer, but rather how well he coordinates several different functions in order to solve problems of daily life. In other words, given this mode of looking at a system, sets of constraining conditions may be regarded as equivalent to models of man or beast which realistically describe human behaviour in a given environment.

We now consider a restricted class of problem-solving tasks—perceptual maze problems which can be formulated as finite multistage decision making problems that may be useful to assess functional brain damage in alcoholics. Dynamic programming, as formulated by Bellman (1968), is an iterative technique for finding optimal decisions for multistage decision making problems. It is desirable to find optimal policies, as without them we cannot satisfactorily analyse the problem solver's behaviour. The maze task to which I am applying some of Bellman's algorithms stems from the work of Alick Elithorn. Maze problems have a long history in psychology, stretching back to the work of Binet and Porteous, and their continued survival seems likely. Yet the maze has been treated in cursory fashion by those who employ it. Benton et al. (1963) produced a set of mazes which were found sensitive to brain damage. In the Elithorn maze (see Figure 1) the subject must pass through as many or as few points as possible. Benton et al. gave the patient a score of one if he correctly solved the maze, and zero if he did not. This scheme has two obvious shortcomings: it makes no distinction between an easy and a difficult maze, and it makes no distinction between a poor solution and one which is good but not complete.

Davies and Davies (1965) present a scoring system based on a graph-theoretical analysis of the maze which has a well-defined mathe-

matical structure. Yet their solution is little better if we wish to progress beyond the stage of ad hoc empiricism. What is required, and what Bellman (1968) provides, is a tool for identifying within a model the parameters which are sensitive to variation due to diffuse cerebral dysfunction.

Using an example of a maze, or more specifically a variable end-point network as provided by Rapoport (1968) we might have, say, that the value of the arc going from P(1, 1) to P(2, 2) is 40, that going from P(1, 1) to P(2, 2) is 49 (see Figure 1). In general, let us associate the symbol A U P(i, j) with the value of the arc going from P(i, j) to P(i, j+1), and similarly A R P(i, j) for P(i, j) to P(i + 1, j).

In the Elithorn maze the task of the subject is to trace a line from vertex (P1, 1) to any node on the line joining P(1, n) and P(n, 1); his path must keep to the lines and must not double back. Thus, we construct things such that the path through the maze or network must be North-East monotonic. A path is N.E. monotonic if, for any of its arcs [P(ij) P(i',j')]

$$\begin{aligned} &\text{either } i < i' \text{ and } j = j' \\ &\quad \text{or } i = i' \text{ and } j < j' \end{aligned}$$

To solve the minimum value problem we define an optimal value function. SP(i,j) = 1, 2 . . ., n at the vertices of P(i,j) of the network under discussion as : SP(i,j) = the value of the (minimum value) admissible path connecting P(i,j) with any vertex on the terminal line.

Since the path from any vertex on the terminal line equals zero we have the following boundary condition:

$$SP(i,j) = O \text{ for } i+j = n+1 \qquad [3]$$

The principle of optimality states that "an optimal policy has the property that: whatever the initial state and initial decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

Applying this principle to our maze problem we can state that:

$$SP(i,j) = \min(W,Z) \qquad [4]$$

$$\begin{aligned} \text{where } W &= A\ P\ P(i,j) + SP((i+1), j) \\ \text{and } Z &= A\ R\ P(i,j) + SP(i, (j+1)) \\ \text{and where } 2 &< 1+ j < n, \end{aligned}$$

and SP(i,j), ARP(i,j), AUP(i,j) are defined as before.

Equations [3] and [4] provide a simple and backward algorithm for finding an admissible path through the variable end-point network. We first compute the minimum path values for the (n-1) one-stage problems.

We next solve the (n-2) two-stage problems in terms of the already known solutions for the (n-1) one-stage problem, and so on.

The forward algorithm follows a similar form, only the indices are changed.

The whole problem becomes more interesting, psychologically speaking, when we consider algorithms which place constraints on the

problem solver. For example, the planning horizon of a person is limited; we may include this in our model. Subjects usually scan from vertex to terminal point so we shall consider here only the forward algorithm. When a forward algorithm F:k is used, subject is assumed to start from vertex P(h) and compute the values of all the paths from P(h) to P(h + k),

$$\text{where } h = i+j = 2,3, \ldots, 2n\text{-}k$$
$$\text{and } k : 1<k<2n\text{-}2 \text{ is a fixed constant}$$

After the values of all paths connecting P(h) and P(h+k) are computed, a path with the minimum value is selected and subject moves one step along this path from P(h) to P(h+1). When in vertex P(h+1) subject computes values of all paths from P(h+1) to P(h+1 +k), picks a path with minimum value and moves along this path to P(h+2), and so on.

We may calculate back algorithms B:k in similar fashion.

Using fixed values of k, Rapoport (1968), has found that these simple models describe *without error* the result of a substantial portion of subjects. Obviously the model may be improved by placing other parameters in the model. k may vary in a non-Gaussian fashion, assuming larger values near the bottom vertex than near the top row. For some other constraints the subject may start at some node P(i,j) moving from this node in two directions—upwards to the top row and downwards to the bottom vertex. Different values of k may be associated with the "forward" and "backward" solutions.

We may extend the use of dynamic programming to situations where the specification of a control does not uniquely determine a solution. Typically, problems in the social sciences will contain variables determined by random or non-Gaussian distribution functions, as well as the usual state-dependent and control-dependent terms. As a result of these stochastic features, a control policy determines a probability measure defined over the space of all possibe solutions, but does not determine which solution actually occurs. We can set up an expected optimal value function, which yields an expected value associated with each control function or policy.

We have been concerned just now with processes which evolve by discrete steps since the concept of a sequence of random variables is intuitively (and computationally) easier to grasp than the concept of a continuous stochastic process. By application of the principle of optimality we can obtain a recurrence relation characterising the optimal expected value function.

As soon as we begin to analyse the concept of information the question of learning arises. If we are given the task of controlling a system about which not everything is known initially, we can try to improve our performance over time by testing and experimenting with different kinds of control actions. We can call a process where both control and learning are involved "adaptive".

Problems just posed represent tremendous scope for computer-human systems research. Already Bellman (1968) is working on a

massive project concerned with the simulation of the psychiatric interview. If we start with the question "is symptom A present" a Yes answer invokes another question pertaining to symptom A, if a negative answer results we ask a question pertaining to symptom B, and so on. The mathematical problem is to do this experimentation and testing in some efficient manner. The overlapping nature of so many symptom groups makes the preceding approach difficult. Nevertheless it is being used successfully in a number of clinics for preliminary screening.

To look at another application of control theory we may take an example from drug research. Once we have constructed a mathematical model that imitates the behaviour of the actual system in sufficient detail, we can begin to study the question of improving the behaviour of the system in many ways, i.e. we can begin to attempt the control of the system. An intermediate problem at present feasible is the determination of the kinds of drugs that ought to be administered for various purposes, dosages and ways in which they should be administered, and in the combinations that give the most efficacious results. In the longer term it may be that little extra effort will be required to guide us in the choice of a new drug family on the basis of results obtained using the previous drugs.

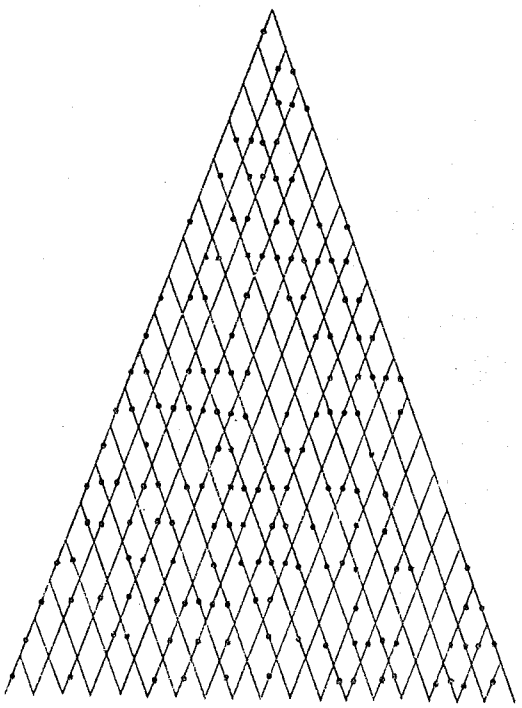

Figure 1. An example of an Elithorn maze.

A passionate skeptic may ask, if we already have a theory of control, why hasn't it found common acceptance? One may accept a comment that new theories are never accepted; their opponents die off. Krantz (1972) has remarked that it is a fact, well known to students of propaganda, that simplification introduced into confusion has high acceptance value. It is not without reason that Krantz is referring to some of the more inbred areas of operant psychology, yet his comment is apposite here. According to Bellman (1968), understanding must involve operational algorithms.

Similarly, Markovian models of learning have features in common with discrete state control systems. Kinsch and Morris (1965) and Waugh and Smith (1962) have shown that in analyses of performance in free-recall situations, a two-stage Markov model gives an acceptable account of the data. The pioneering work done in psychology using dynamic programming techniques shows that these theories of behaviour are testable, and can be modified to fit the data.

The work described in this paper was supported by a grant, from the War Pensions Medical Research Trust Fund Board, held by Professor R. A. M. Gregson.

REFERENCES

Bellman, R. *Some vistas of modern mathematics.* University of Kentucky Press, 1968.

Benton, A. L., Elithorn, A., Fogel, M. L., and Kerr, M. A perceptual maze test sensitive to brain damage. *Journal of Neurological and Neuro-surgical Psychiatry,* 1963, 26, 540-544.

Cornfield, J. The Frequency Theory of Probability, Bayes Theorem and sequential clinical trials. In D. L. Meyer and R. O. Collier, *Bayesian statistics.* Illinois Peacock, 1970.

Davies, A. D. M. and Davies, M. G. The difficulty and graded scoring of Elithorn's perceptual maze test. *British Journal of Psychology,* 1965, 56, 295-302.

De Groot, M. M. *Optimal statistical decisions.* McGraw-Hill, 1970.

Dreyfus, S. E. *Dynamic programming and the calculus of variations.* New York: Academic Press, 1965.

Krantz, D. L. Schools and systems: the mutual isolation of operant and non-operant psychology as a case study. *Journal of the History of the Behavioral Sciences,* 1972, 8, 86-102.

Kinsch, W. and Morris, C. J. Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology,* 1965, 69, 200-206.

Rapoport, A. Optimal and suboptimal decisions in perceptual problem solving tasks. *L. L. Thurstone Psychometric Laboratory, University of North Carolina Research Report,* No. 64, 1968.

Toda, M. The design of a fungus eater: a model of human behaviour in an unsophisticated environment. *Behavioral Science,* 1962, 7, 164-183.

Waugh, N. C. and Smith, J .E. K. A stochastic model for free recall. *Psychometrika,* 1962, 141-154.