Table 3

Adjustment of multipliers to a sliding origin

After the *j*th play μ is calculated as

$$\frac{1-D}{D-D^{(j+2)}}\sum_{i=0}^{j} D^{(j\ \ i+)}V_i,$$

where V_i is the outcome value of the *i*th play and V_0 is set equal to 0 (value of a win is +1, of a draw is 0, and of a defeat is −1). D is the decay factor and M_n is the unadjusted multiplier for the *n*th stage of the game (see text).

OUTCOME	REINFORCEMENT
Won	$R_n = M_n^{-\mu+1}$
Drawn	$R_n = M_n^{-\mu}$
Lost	$R_n = M_n^{-\mu-1}$

allowed to learn. After a few hundred games the two sides were both producing near-expert play.

Improvements to the program

These results are only preliminary. The program has now been modified so that the value of an outcome is assessed against the average outcome of past plays, instead of remaining fixed. It seems obvious that a draw, for example, should be rewarded when the usual outcome has been defeat, and punished when the usual outcome has been victory. Similar considerations apply to the values of winning and losing outcomes. The method which has been adopted is the following.

The value of a win is rated at +1, that of a draw at 0 and that of a defeat at −1, and a weighted average, μ, of past outcome values is formed using as weight a decay factor D ($0 < D \leqslant 1$). Thus the weight of the last outcome is D, that of the penultimate outcome is D^2, that of the antepenultimate outcome is D^3, and so on. The smaller the value chosen for D, the more weight is given to recent experience; as D approaches unity, increasing weight is given to the experience of the more remote past. In theory, a running calculation is made to evaluate μ after each play, and this is used to adjust the multipliers as shown in Table 3. The implementation in the current version of the program does not actually attain this ideal, but makes an approximation. The decay factor is only applied to the average of each set of one hundred plays.

Our model of trial-and-error learning is thus based on three adjustable parameters, A, B and D (see Fig. 6 and Table 3). The next paper of this series will describe the effects upon learning performance which result from the systematic variation of these parameters.

Acknowledgements

The work described was supported by a Grant-in-Aid from the Royal Society. My thanks are also due to Messrs. Bruce Peebles and Co., Edinburgh, and to Mr. W. A. Sharpley personally, for making computing facilities available to me.

Reference

MICHIE, D. (1961). "Trial and Error," *Science Survey*, 1961, Harmondsworth: Penguin, Part 2, pp. 129–145.

Correspondence

To the Editor,
The Computer Journal.

Dear Sir,

"*Direct coding of English language names*", *The Computer Journal, Vol.* 6, *No.* 2 (*July*), *p.* 113

Surely the duplication in book titles tends to occur at the *beginning*. Could a solution be found for a short unambiguous code in referring to the *last* word, say the first and third, or better still the ultimate and antepenultimate?

e.g.	Selections from Borrow		SLWR
	Selections from Byron		SLNR
	Short History . . . etc. . . .	Augurelius	SOSI
	Short History . . . etc. . . .	Augustus	SOST

Yours faithfully,

E. R. KERMODE.