Experiments on the mechanization of game-learning

By Donald Michie

Donald Michie

- 1923-2007
- biologist
- Nature: *Father of artificial intelligence in Britain*

http://www.aiai.ed.ac.uk/~dm/donald-michie-2003.jpg

receiving his honorary degree from Stirling University in 2003

Trial and Error

Can computers think?

We might even decide to define 'thinking' to include the subjective experiences of the thinker; it would then follow automatically that insentient beings, which might be held to include machines, cannot think.

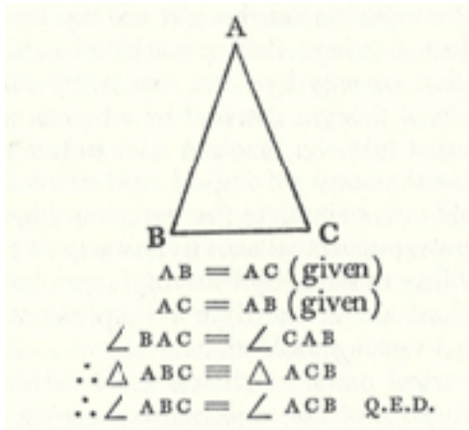
It will therefore not be through perversity, but through need, if in describing mechanical processes I intermittently borrow words from the vocabulary of human or animal psychology.

Human Intellectual Activity

Originality and Ability to learn

Originality

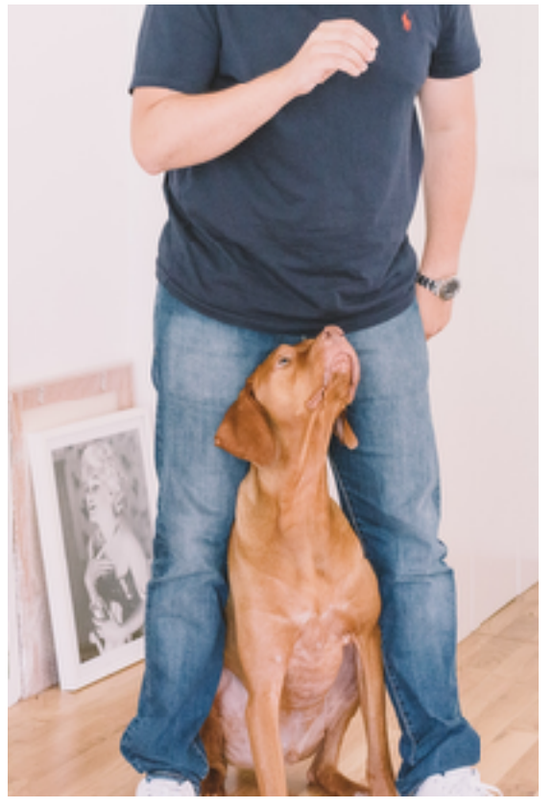
- Marvin Minsky shows: computer can find new proof for theorem of Euclid



No biologist in his senses would look at a modern aeroplane and conclude that birds, despite appear- ances,must work on a jet-propelled fixed-wing principle, but the temptation sometimes presents itself in more subtle guises.

All that we have a right to expect from a model is that it may deepen our understanding of the matrix of phy- sical laws within which both the model and the biological system have to work

Components of Trial and Error Learning



Classification of the stimulus

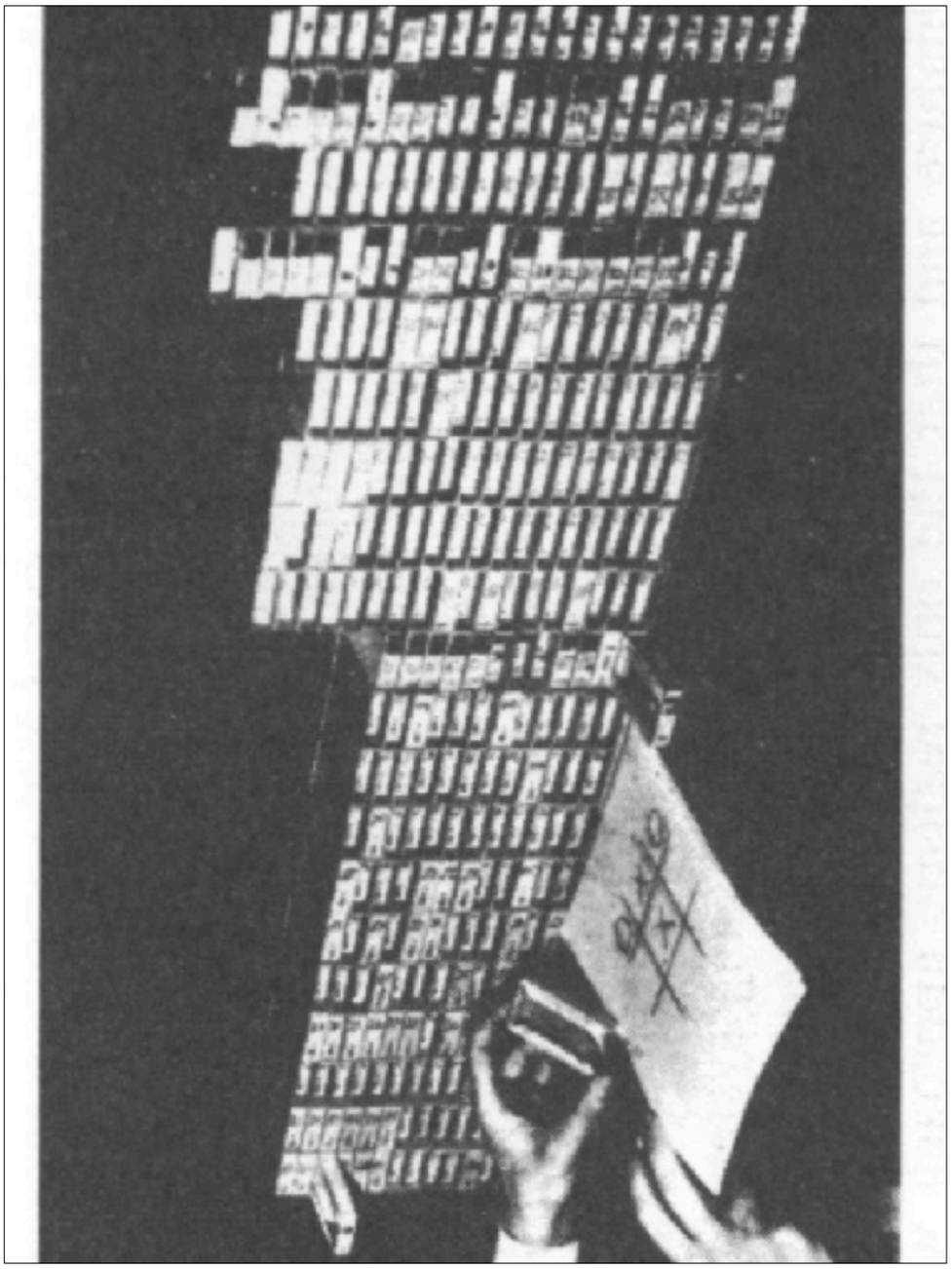
Reinforcement of the response

which will otherwise have to be spoon-fed with texts laboriously punched on to teleprint tape by human typists.

Reinforcement

- Events have and outcome
- outcome has value
- value: degree of pleasure or displeasure associated
- positive value: chance of outcome increased

To learn: number of discrete situations to learn must be sufficiently small for all of them to be separately enumerated



one per square

only legal moves can be made. But: would learn that anyways

Symmetry.

First move: only three boxes. Side, corner, middle

Reinforcement

- Machine lost: Take bead from each open box
- Machine won or drew: place extra bead(s) into each open box

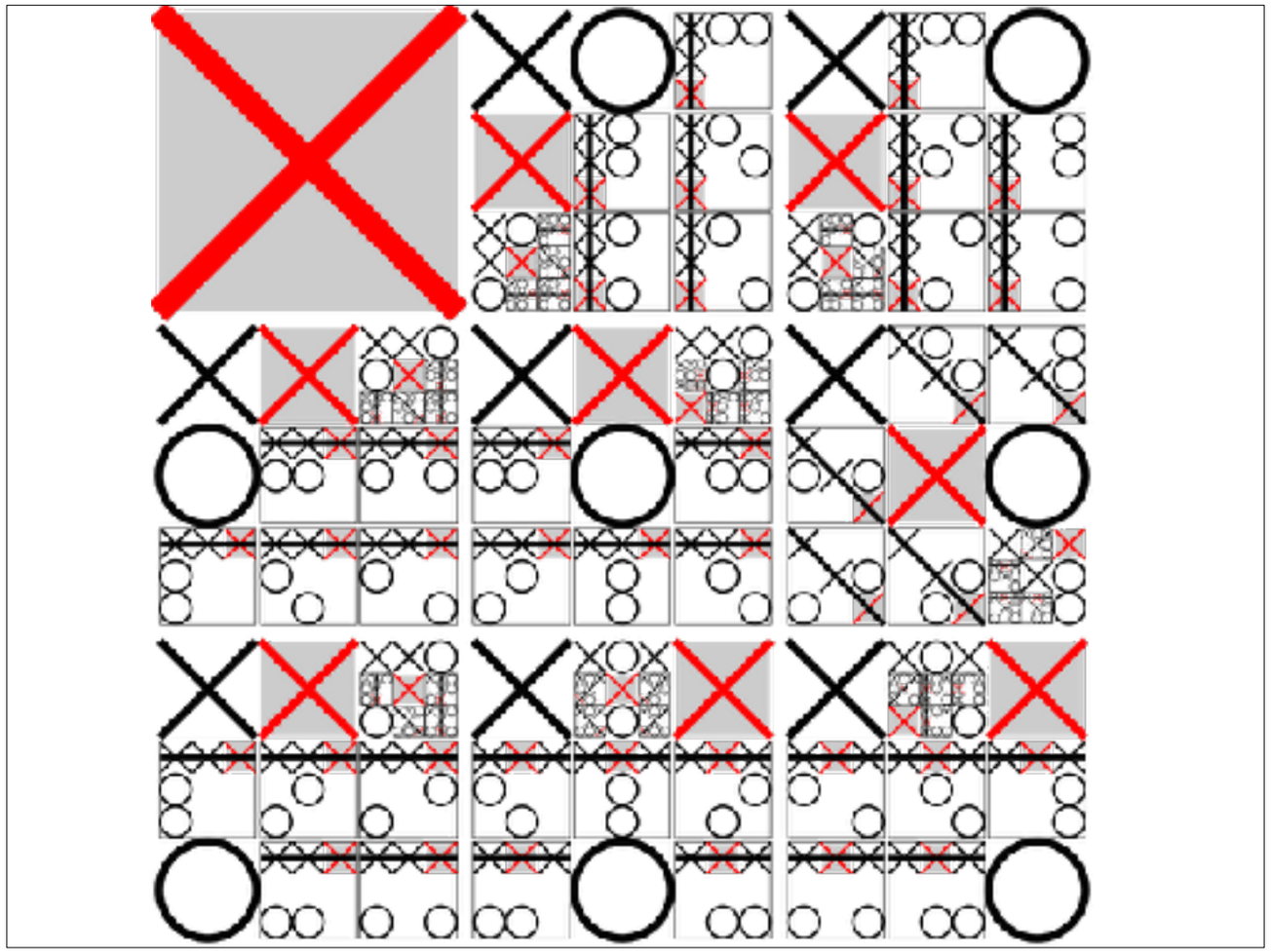
Initial distribution

stage	machines's move	number of replicates
1	1st	4
3	2nd	3
5	3rd	2
7	4th	1

Boxes can run out: Machine gives up. This is okay

If first box runs out: Machine refuses to play

"the time came for the machine to be challenged by its inventor"



Best strategy: Impossible to win

“One might therefore think, assuming that its human opponent would adopt best strategy, that the question of rewarding MENACE for victories would not arise. But in practice the machine quickly found a safe drawing line of play against best strategy, so that its human opponent had to resort to unsound variations, risking machine victories in the hope of trapping it into a more than compensating number of defeats. This possibility had been foreseen (although not the speed with which it matured) and the bonus for a win was fixed at three beads added to each open box. ”

220 plays
2 eight hour sessions
after 150 plays: can only draw
gave up after 8/10 loss
“it is likely, however that my judgement was sometimes impaired by fatigue”

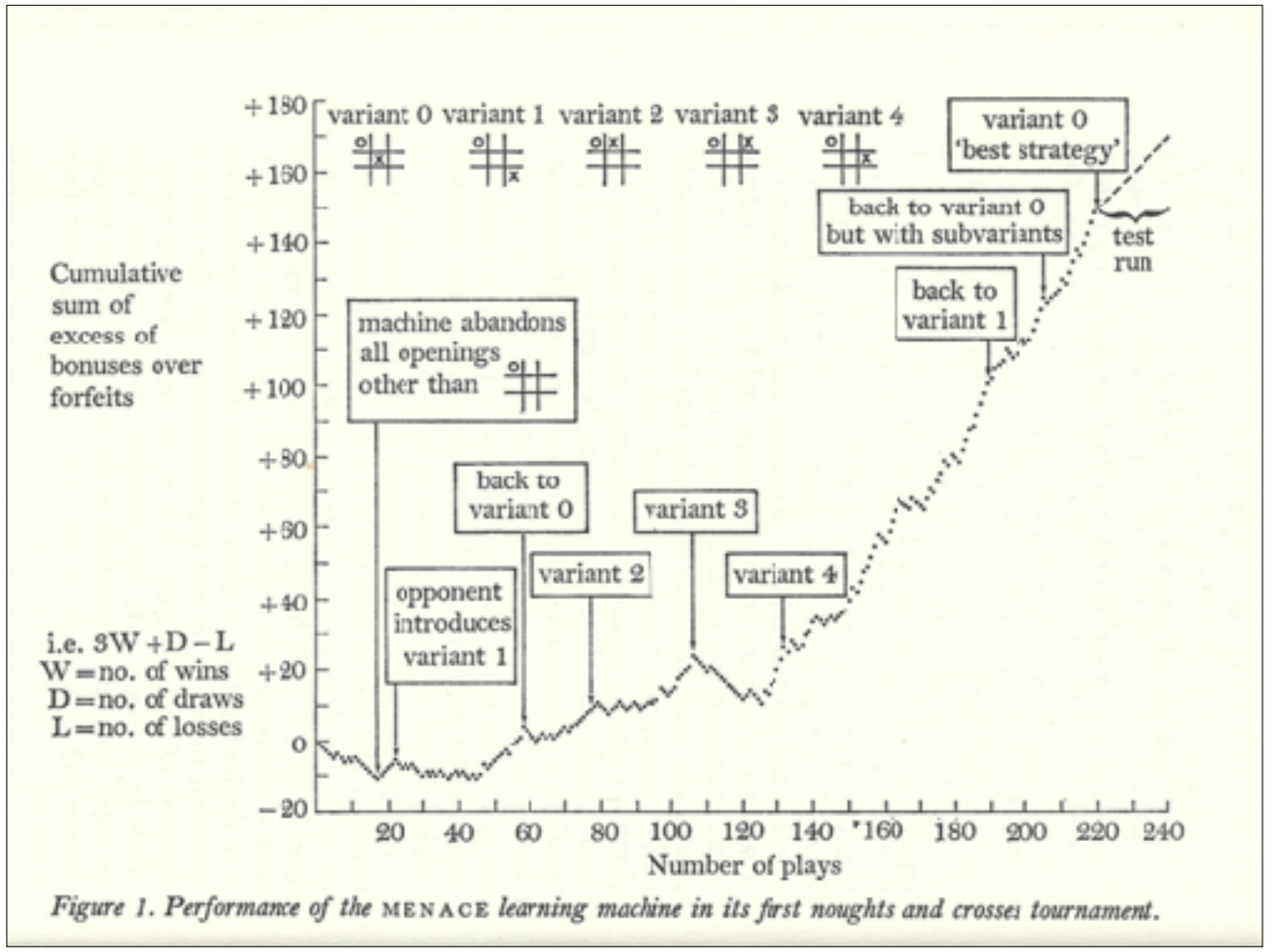


Figure 1. Performance of the MENACE *learning machine in its first noughts and crosses tournament.*

45°: average drawing result: 1 play, 1 bead more

Experiments on the mechanization of game learning

Ferranti Pegasus 2 (14 sold)

plays both players

1 game per second

random game: beginner wins about 2 in 3

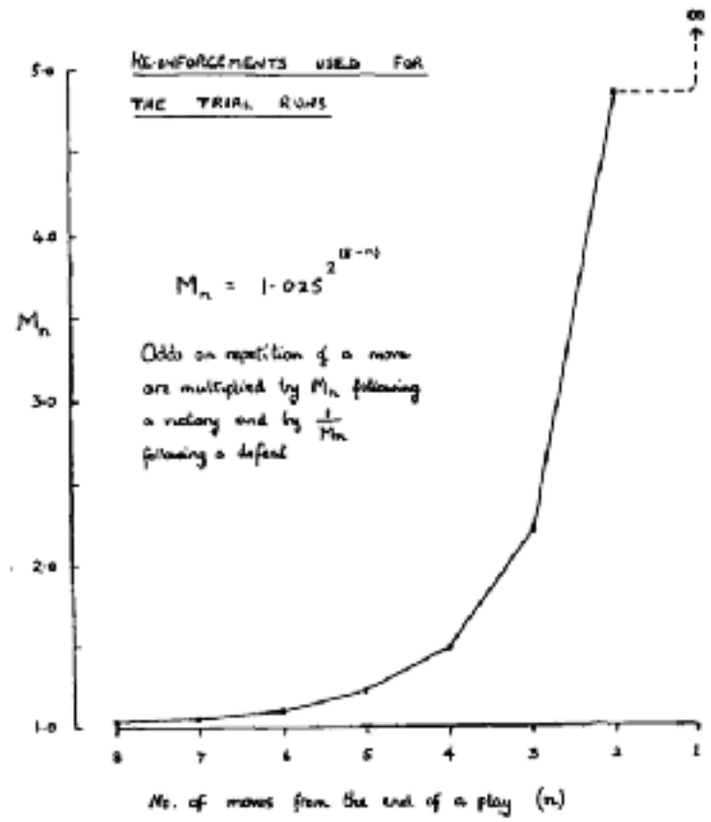


Fig. 6.—The multipliers used for reinforcement in the trial runs of Figs. 6–8. A special case is shown of the general form $M_n = A^{B^{(8-n)}}$

Probability adjustment

- Probabilities are stored as odds ($\frac{p}{1-p}$)
- with $M_n = 2$ and $p_1 = \frac{2}{5}, p_2 = \frac{3}{5} \Rightarrow p_1 = \frac{4}{7}, p_2 = \frac{3}{7}$

- third-last move.
- only two choices left
- 2:3 - multiplier 2 - after: 4:3

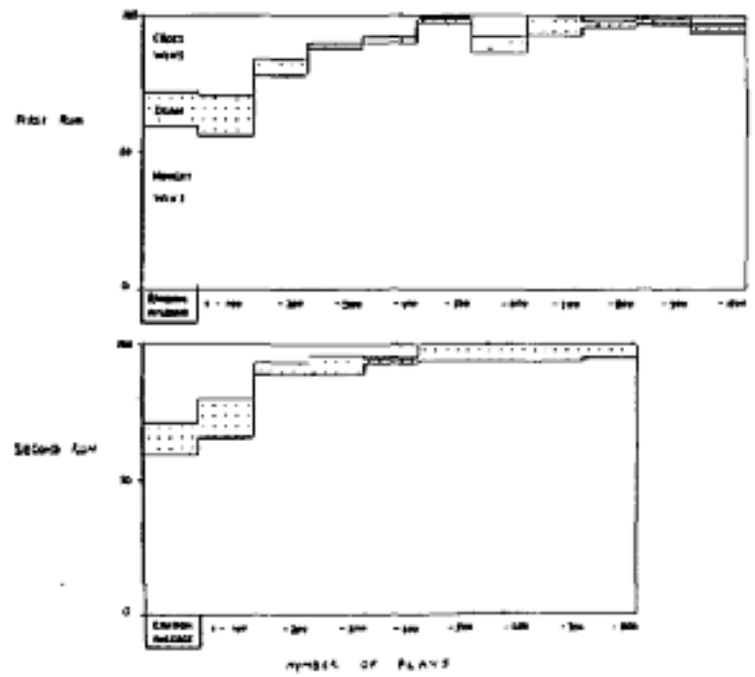


Fig. 7.—Trial runs with the computer program. Nought (the opening player) is learning, while Cross is playing at random throughout. The left-hand block shows the average results when both sides play at random

assumption: reinforcements too strong -> premature conclusions

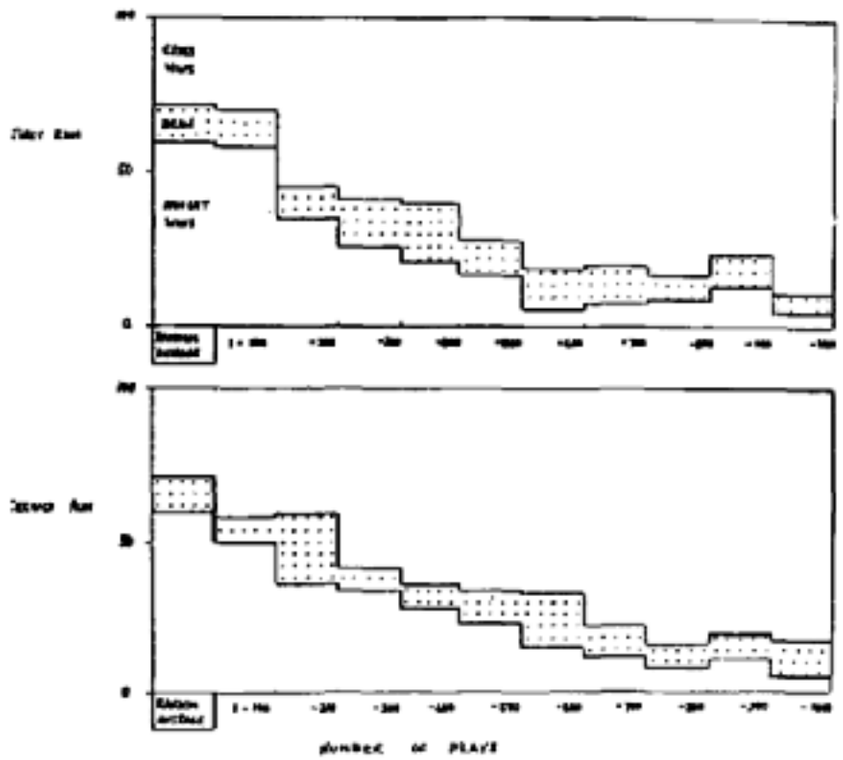


Fig. 8.—Cross is learning, Nought playing at random

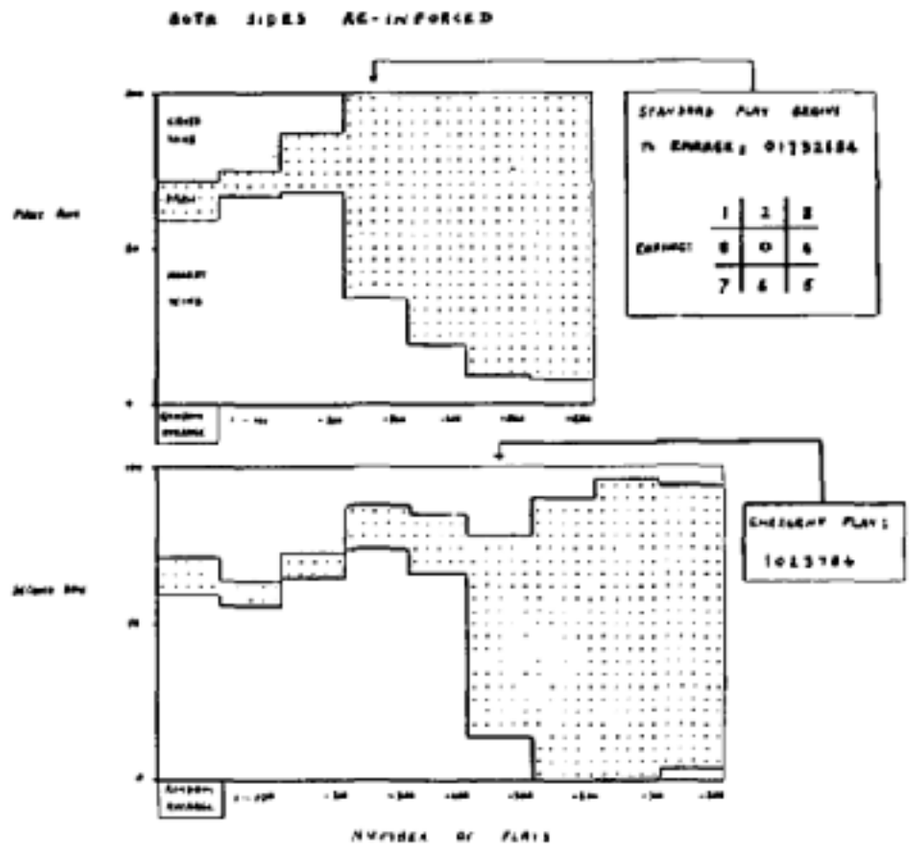


Fig. 9.—Both sides are learning

Improvements

- Value of outcome is assessed against average outcome of past plays
  - Win: +1
  - Draw: 0
  - Defeat: -1
- Decay factor: *D (0 < D ⩽ 1)*

Adjustment of multipliers to a sliding origin

After the jth play μ is calculated as

$$\frac{1-D}{D-D^{j+2}}\sum_{i=0}^{j} D^{j-i}W_i$$

where W_i is the outcome value of the ith play and W_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 nth stage of the game (see text).

OUTCOME	RE-ENGAGEMENT
Won	$R_n = M_n^{n+1}$
Drawn	$R_n = M_n^{n}$
Lost	$R_n = M_n^{n-1}$

This gives parameters A, B, D (M = A**B**(8-n))

outlook: Effect of variation will be topic of next paper

Try it

http://www.mscroggs.co.uk/menace/