MULTISTAGE ACQUISITION OF INTELLIGENT BEHAVIOUR*

B.D. Josephson and H.M. Hauser**

Cavendish Laboratory, Madingley Road, University of Cambridge, Cambridge, CB3 OHE, U.K.
(Received April 18, 1980)

Abstract

Human skills are acquired not by a single uniform process, but in a series of stages, as Piaget has shown. We have investigated such a sequential process by taking as an illustrative example the game of table tennis. The aims in each stage of learning are qualitatively different, and we show in detail how knowledge gained during one stage provides essential information for subsequent stages. Conclusions are drawn which may be important for artificial intelligence work generally. The question of practical implementation of a system such as discussed is considered briefly.

1   INTRODUCTION

In this paper we are concerned with the question of how a human being *becomes* intelligent. This is a different aim from that of most artificial intelligence work, which is concerned mainly with a description of intelligent behaviour, usually in the form of a computer program. We are concerned here with how the finished product comes to be produced, more than with the finished product itself.

The key to our approach is the discovery by certain psychologists, most notably Piaget [1], that skill acquisition occurs in a number of discrete stages; for example, there are six stages in the development of sensory-motor skills. The picture of man's development in stages as indicated by Piaget can, perhaps, be usefully compared with that of a program written in a language such as Hewitt's Planner[2]. In the usual AI context, the latter would contain information such as "if you want to open a door, rotate the doorknob and move it towards or away from you". Man's general evolutionary program, on the other hand, would contain information such as "if you want to be able to achieve things in life, learn to talk" and "in order to learn to talk, you must learn to distinguish the sounds of speech you hear, and then to copy the sequences that other people use".

Similar evolutionary principles apply to particular segments of life, such as the one considered here, that of playing table tennis. A player of this game must learn to master skills such as being able to hit the ball and to be able to make it go in the direction he wants, before he can have any hope of learning the more subtle skills involved in becoming an expert player. A beginner cannot become an expert by imitating the actions and strategies of an expert; his ambitions in the first instance must be much more limited. This much is probably obvious, but what we may hope to be able to do is to understand the deeper reasons behind the subdivision of skill acquisition into stages. The analyses in sections 2 and 3 suggest

that the basic purpose of the subdivision is to render an impossibly complex task feasible. What we find is that each stage of skill acquisition provides knowledge which in a sometimes quite subtle way simplifies the learning task involved in the next stage. The order of two stages in the sequence of skill acquisition cannot profitably be reversed.

These conclusions can be made with a fair degree of confidence in the case of sensory-motor skills, but may be of importance in the field of artificial intelligence generally. It may be a fruitless task to attempt to make a computer program solve difficult problems by feeding it with a diet of difficult problems only. It may be necessary, as it is with human beings, to present it with a carefully graded series of problems with no large gaps in difficulty from the very easy to the very difficult. Scene analysis is discussed in this light in section 4. If this judgement is correct, some reorientation of goals within AI may be called for.

While it is not the intention of the present paper to provide detailed models for implementing the processes discussed, it has been felt desirable to indicate general mechanisms which might allow such implementation. This task is undertaken in section 5.

2.   THE STAGES OF SENSORY-MOTOR DEVELOPMENT

We shall begin by giving a very brief description of what is accomplished during a number of stages in the development of the skill of playing table tennis.

In stage 1 the player learns a general description of the game, from watching the game and from verbal description. He learns (a) what to expect to happen in a given situation and (b) what he is expected to do: the actual actions and what they are intended to achieve. The latter is obviously determined by the rules of the game.

In stage 2 the player learns the basic skill of being able to hit the ball, and in stage 3 to be able to control its direction. In stage 4 he learns to optimise a given stroke by choice of body position and orientation, and in stage 5 to choose in advance the most suitable type of stroke for what he wants to achieve. Finally in stage 6 the player learns to direct his actions to achieve the optimum effect (in terms of making things difficult for the opponent, for example).

It is instructive to examine these stages from the viewpoint of operant conditioning theory. This states that those actions which in a given situation lead to some reward or reinforcement tend to be repeated when the same situation occurs again.[1] In the present instance it is easy to see that the general *types* of reinforcement are different in the different stages, so that the learning algorithm can operate by selecting different types of event to be reinforcing at different times (see section 5). Specifically, the rewards are in stage 1, being able to predict and/or understand what is seen to happen, and in stages 2 and 3 respectively being able to hit the ball and being able to match its actual direction with the direction intended. The rewards in stages 4 and 5 are rather more subtle; in stage 4 it is probably the naturalness and ease of the stroke and the actions which precede it, as will be explained later, and in stage 5 the degree of success in difficult situations (since in a difficult situation only the best choice of stroke is likely to be successful). In stage 6 the reward is the degree of difficulty experienced by the opponent.

The order of the stages given above cannot profitably be reversed, as knowledge gained in one stage is needed for the next. This point will be understood more clearly after the discussion in the next section, but may be discussed in qualitative terms now. We take stages 4 and 5 to illustrate the point. In stage 5 what he learns is the best stroke to choose on the basis of its success in difficult situations (e.g. whether it actually lands on the table on the right side of the net or not). But the success of a stroke is not a particularly good measure of the correctness of choice unless the stroke is carried out reasonably proficiently, and this is learnt in stage 4 (more specifically, going through stage 5 without

going through stage 4 first leads to the adoption of "bad habits"). Such a
problem does not arise when the stage are gone through in the correct
order, as a good style can perfectly well be acquired even if a player
sometimes chooses an inappropriate stroke.

3 CONTROL OF THE INFORMATION EXPLOSION

The scheme of sequential development outlined here can be looked at in
another way, namely as a way to prevent the amount of knowledge to be
learnt and of information to be processed from becoming too great. This
complexity has two components, the amount of input information (the number
of possible ball trajectories, for example) and the large number of actions
which might possibly be considered in response to a given situation. Let
us now consider some particular illustrations of this point.

We need not dwell on the application of this principle to stage 1;
obviously this pre-programming means that the subsequent activities are
not a matter of blind trial and error. In stage 2 the obvious result of
being able to hit the ball is accompanied by that of being able to
represent the complex visual information in a form which is particularly
useful for subsequent stages. What the player learns during stage 2 is
the *configuration of his body* at the moment of impact. In Piaget's
terms, the environment is mapped on to an action; equivalently, an
important component of the total visual information has been abstracted
from it and can be used instead of it in later stages. This component
contains no information about the *direction* of the ball, but the
requirements for the latter information are not as stringent as those for
the precise position.

How can stage 2 knowledge be acquired? An important consideration seems to
be that the unambitious nature of the goal at this stage allows a large
degree of *uniformity* in the response. Since a two-parameter family of
trajectories can fill a region of space, the player can achieve his goal by
specifying only two parameters to control his arm trajectory (this will
allow him to hit a desired point of the ball trajectory; to ensure correct
timing he must also learn the appropriate visual cue to start the forward
movement of his arm). These parameters might be used simply to determine
the position to which his arm moves back before beginning his forward
stroke, and they might not be used to control the forward stroke at all.
From his successful shots the player can learn to associate with the visual
information the two parameters used to set up the trajectory and a third
parameter associated with the arm position at the moment of impact (the
latter being indicated by auditory, tactile and kinaesthetic cues). These
three parameters indicate to a high degree of precision the position of the
ball at impact and represent the required abstraction of information.

Stage 3 is not concerned with dealing with the information explosion, but
with producing a flexible response. It rather *creates* an information
explosion, because of the number of adjustment parameters that have to be
learnt: adjustments in stroke direction to allow for different directions
of the incoming ball and different target directions, and the arm position
adjustments required to ensure that contact with the ball can still be made
in spite of the alteration in stroke direction. It is unlikely that
anything better can be learnt than linear adjustments, valid within limited
regions surrounding a set of preferred arm trajectories. It can be seen as
the aim of stage 4 (co-ordination of body movements and arm movements) to
extend the viability of such schemes, by permitting a body movement to
bring the required arm movement into the optimum region. As suggested in
the preceding section, this might be done by having a possibly innate
system to specify certain arm movements as more natural or comfortable.
During stage 3 a player would learn that certain visual information would
correlate with an uncomfortable stroke, and during stage 4 he would use
such information as input to learn how to move his body to ensure a stroke
of maximum ease and comfort. Having done this, he would have to a large
extent overcome the problem associated with stage 3, that of the
adjustments being satisfactory over only a limited region.

Skill acquisition up to and including stage 4 is concerned with perfecting

a given kind of action, but after this it involves the selection of styles from a discrete set, in a way very similar to that involved in biological evolution of species. Now the player himself is the agent of natural selection, selecting on the basis of his past success. His selection at any given moment consists of a set of binary decisions, such as forehand or backhand, topspin or chop, maximum power or maximum precision and so on. As he accumulates experience his binary decisions become more clearcut, and the problem of keeping track of the information explosion lessens as only a few more successful strategies remain. At the same time, however, the player introduces occasionally 'mutations' or slight variations on old styles, in an attempt to give himself an even more useful selection of styles to choose from.

4 MORALS FOR ARTIFICIAL INTELLIGENCE

Many of the points discussed above may seem very obvious to the reader. it must be realised, however, that the principles revealed are ones which to a very large extent are not used in Artificial Intelligence (AI) programs. For example, the concepts of learning to achieve simple goals not directly related to ultimate goals, of learning good values of parameters by means of considerable trial and error, and evolutionary principles analogous to those operating during biological evolution do not figure very prominently in most of AI, and yet in problems like that considered here they seem to be very useful. Scene analysis according to Piagetian concepts is very different from that according to the concepts of AI. For example, a child would first learn to discriminate objects from the backgrounds on the basis of cues such as motion parallax and the fact that an arm movement sufficient to reach an object may often not suffice for reaching the background, even if it is considerably extended. His future ability to recognise objects will be on the similarity of some characteristic feature to one of an object already examined, rather than on some absolute ability to analyse any collection of objects into components. Discriminations will be made on the basis of the usefulness of so doing (for example differences in colour which correspond to differences in taste will be noted), rather than in accordance with any absolute classification scheme. Visual cues used in AI programs such as the configuration of edges at a vertex almost certainly are used but only because they have been correlated in the past with successful figure-ground discriminations. In conclusion, it may be suggested that a close study of observations of the type pioneered by Piaget, coupled with careful analysis of their significance, might be extremely valuable to AI generally.[2]

5 SOME PROBLEMS OF IMPLEMENTATION

The remarks in this section are not in any sense intended to be a precise theory, and consist only of rough suggestions as to how some of the variety of operations of information processing and knowledge acquisition discussed in sections 2 and 3 might be implemented in a practical system.

The operant conditioning concept requires that similar actions occur in similar circumstances if these actions have been reinforced previously. How can the concept of similarity be represented in a useful way? It seems that quite often important concepts to which the idea of similarity is applicable vary over a two-dimensional space. Examples are the two-parameter family of trajectories described in section 3, and the two-parameter specification of a direction. Other examples, involving perception, are colour (hue + saturation) and the sounds of the vowels characteristic of a language. Since in the physical nervous system there are very often two dimensional arrays of standard neuronal circuits it is very likely that this type of specification of information in terms of two-parameter sets is implemented as *spatial localisation* of the relevant nervous system activity. According to such a model, learning to repeat a given action, such as that of moving the arm back to a given position (as in stage 2), in response to a particular cue, is in principle a matter of learning to produce activity in a specified region of the nervous system in response to the cue. This could be achieved by models such as that of Wilshaw et al.[4].

In other situations, such as the learning of fine adjustments involved in stage 3, a different mechanism is probably involved. The input signal which indicates the degree of adjustment required may be assumed to alter the *number* of excited neurons belonging to a particular population, and the learning problem reduces to that of adjusting the average *output per neuron* till the correct proportional adjustment is made.

It is also necessary that the information fed in to control the response should be suitably specific, i.e. that the input should not change much if the situation is similar as far as the action is concerned. This requires both filtering and preliminary interpretation of the raw data from the sense receptors. It can be seen that previous experience may be of key importance for this task, a concept very much in line with the general them of this paper. For example, exposure of an individual to white objects occupying large areas of the visual field will excite a characteristic population of neurons which are just those which will respond to the ball in a game of table tennis, wherever the ball may be located in the visual field. Again, experience with reaching out to and walking towards perceived objects during childhood will have enabled the player to interpret a particular visual stimulus as an object at a particular distance, in a manner similar to that described for stage 2 learning.

Finally, one can ask what mechanisms might be involved in causing a player to advance in turn through the various stages when he is ready for them. A simple answer is to suppose that reinforcements are ordered in terms of quality. When a player finds he is successful at a given quality level (i.e. few negative reinforcements occur at that level) he is no longer reinforced at that level and seeks positive reinforcement at a higher level. On the other hand, when he performs badly at one level (too much negative reinforcement) he lowers his aspirations and is content to achieve lower level reinforcement.

REFERENCES

[1] H. Ginsburg and S. Opper, Piaget's Theory of Intellectual Development (Prentice-Hall, Englewood Cliffs, N.J., 1969).
[2] C. Hewitt, PhD. Thesis (AI-TR-258, Massachusetts Institute of Technology).
[3] A.R. Luria, The Working Brain (Penguin, London 1973).
[4] D.J. Wilshaw, O.P. Buneman and H.C. Longuet-Higgins, "Non-Holographic associative Memory", Nature 222, 960 (1969).

FOOTNOTES

[1]It is reasonable to assume that near misses are equally good in the learning situation, since quite often when a near miss has occurred the player can infer quite accurately what action would have led to reinforcement.

[2]It may be worth drawing attention to the work of Luria[3] in which considerations similar to those used here are combined with detailed evidence from neuropsychology.