BIOLOGICAL LEARNING AND ARTIFICIAL INTELLIGENCE

Christian Balkenius

Lund University Cognitive Science
Kungshuset, Lundagård, S–222 22 LUND, Sweden
Christian.Balkenius@fil.lu.se

Abstract: It was once taken for granted that learning in animals and man could be explained with a simple set of general learning rules, but over the last hundred years, a substantial amount of evidence has been accumulated that points in a quite different direction. In animal learning theory, the laws of learning are no longer considered general. Instead, it has been necessary to explain behaviour in terms of a large set of interacting learning mechanisms and innate behaviours. Artificial intelligence is now on the edge of making the transition from general theories to a view of intelligence that is based on an amalgamate of interacting systems. In the light of the evidence from animal learning theory, such a transition is to be highly desired.

1. INTRODUCTION

For many years, researchers within both animal learning theory and artificial intelligence have been searching for the general laws of learning. I want to propose that such laws cannot be found for the simple reason that they do not exist. Below, I will give a number of examples of classical experiments that have shown that a number of mechanisms are involved in learning, none of which is general enough to suffice in all situations. Any attempt to construct artificial intelligence based on one or a few simple principles is thus bound to fail.

The classical strategy in artificial intelligence has been either to depend on an axiomatic system as in the logical tradition (*cf.* Charniak and McDermott 1985) or to base all intelligence on a simple principle such as chunking (Newell 1990). In both cases, the problem of intelligence is reduced to that of searching (*cf.* Brooks 1991). The problem of search control, is however, still mainly unsolved and as I will argue below, will remain so unless artificial intelligence makes the transition into a more diversified view of intelligence.

Before I start the presentation, let us first see what I mean by learning and intelligence. Both terms are, of course, exceedingly vague and I will make no attempt to change that situation. We have, nevertheless, some intuitive appreciation for the meaning of the two concepts and no harm can come from subjecting them to closer examination[1].

Konrad Lorenz defined learning as *adaptive changes of behaviour* and that is indeed the reason for its existence in animals and man (Lorenz 1977). However, it may be too restricted to exclude behavioural changes that are not adaptive. There are in practice, many behavioural changes that we would like to call learning although they are not at all adaptive. We should not forget, however, that these instances of learning are more or less parasitic on an ability that was originally constructed to control adaptive changes. Hence, it seems reasonable to consider learning as a change in behaviour that is more likely than not to be adaptive.

We now turn to the concept of intelligence. Behaviour is usually considered intelligent when it can be seen to be adaptive. An animal is considered intelligent when we can see how its behaviour fulfils its present or future needs. A squirrel that hides nuts in apparent anticipation of the winter is thought of as more intelligent than a lemming that throws itself over a cliff. But when we learn that the squirrel will continue to collect nuts even when it has hidden infinitely more than it can possibly eat over winter,

---

[1] It is with some hesitation I introduce a concept such as intelligence. While it was once a required ingredient of any text on learning, today it is more often than not considered a mortal sin.

we begin to question its intelligence. Eventually, we hear that it does not even remember where it has hidden its winter supply, and the case for squirrel intelligence is settled.

This example shows that we call behaviour intelligence only when *we* see how that behaviour is adaptive for the animal. This is precisely the idea that "intelligence is in the eyes of the observer" (Brooks 1991). We should not, however, be fooled to believe that intelligence is *only* in our eyes. If we change the environment of the animal in such a way that its initial behaviour is no longer adaptive, we can make an interesting observation. If the animal persists in its original behaviour, we no longer consider it intelligent, but if it, on the other hand, changes its behaviour to adapt it to the new circumstances, we will still think of it as intelligent in some sense. This, I think, makes intelligence equivalent to the capacity of learning.

2. THE LEGACY OF BEHAVIOURISM

During the reign of behaviourism it was habitually taken for granted that all behaviour could be explained in terms of stimulus–response (S–R) associations. Based on this belief, innumerable experiments were conducted with one singe goal in mind: to establish the general rule for S–R formation. Once this rule was discovered, we would know everything there was to know about learning.

Following this line of though, it seemed reasonable to simplify the learning situation as much as possible until only the essential core of the task was left. In the early experiments, researchers were using a small copy of the garden maze at Hampton Court for their animals (Small 1901). This maze turned out to be much too complex[2] and as time went on the mazes became simpler and simpler until the development culminated in the ingenious Skinner box. This device was entirely devoid of any behavioural possibilities except for bar pressing. While the animal in the Hampton Court Maze could perform a large number of actions, the rat in the Skinner box could do only one of two things; either it could press a lever and receive food or it could refrain from doing so.
One may object that there are many ways to press the lever and even more ways to refrain, but all these cases were conveniently lumped together using operational definitions of the two cases. It was the movement of the lever that counted as a response, not the movement of the animal. Hence the name *operant* learning procedure.

[2]To complex for the researchers, that is, the animals did not have any troubles. In fact, when rats are given a choice, they prefer to explore complex mazes instead of simple ones.

Based on the fundamental belief that all behaviour in all species could be explained in terms S–R associations it was entirely immaterial whether we would study rats in the Skinner box or humans learning university mathematics. The process involved would be the same. Of course, it was much more practical to study rats in the laboratory and that is how the research proceeded.

One may ask if the animals had any choice other than to learn an S–R association? What else was there to learn? The experimentalists had removed all other possibilities for learning based on the presupposition that they did not exist. Consequently, they had eliminated all possibilities of disproving their underlying assumption. Years of effort were devoted to the simplest form of learning conceivable. It is the irony of the whole approach that we still, almost 100 years after Pavlov's and Thorndike's initial experiments, do not know exactly what rules govern the formation of the supposed S–R association.

3. THERE IS NOTHING GENERAL ABOUT LEARNING

What would happen if we arranged for other types of learning than pure stimulus–response formation? What if we construct tasks where learning of simple associations does not suffice? Let us look at some experiments.

One of the very first experiments to question the view that *responses* were learned was conducted by Macfarlane in 1930. He trained rats to swim in a maze in order to obtain food placed on the goal platform. When the rats had learned their way in the maze, it was drained of water and the rats were again placed in the start box. It turned out that they could still approach the goal with almost no errors even though they were now running instead of swimming.

Whatever they had learned, it could not have been the response of performing some specific swimming motion associated with the stimuli at each place in the maze. According to Tolman, the rats had not learned a series of responses but instead the spatial layout of the maze. This 'cognitive map' could then be used to get from the start to the goal in any of a number of ways. While this experiment certainly shows that something more abstract than a S–R association was learned, we cannot resolve the question as to whether it is anything like a cognitive map or not. For this, we need more evidence.

Another of Macfarlane's experiments was again supposed to show that animals learn a map of the maze and not a response chain. In this experiment, animals were trained to find a goal box in a simple T maze. Once the rats had learned the place of the food,

the maze was turned 180° and the food removed as shown in figure 1. As a result, the arms of the maze were interchanged. If the rats had learned to make the response of turning right at the choice point, they would continue to do so even after the maze was turned. If they, on the other hand, had learned the spatial location of the food they would now turn the left. And so they did. Again it could not have been the response that had been learned.

It has later been shown that under some circumstances the rats will continue to turn right. Here, we will not consider what determines which strategy the animals will use. The important observation is that in *some* cases, they will use a place strategy.
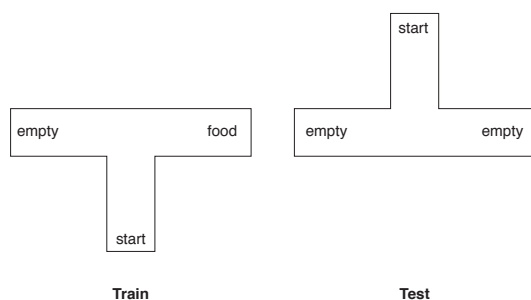

Figure 1. Mazes for a place-learning experiment

Mackintosh (1983) distinguishes between three types of possible learning mechanisms in simple T-mazes. If the two arms of the maze are physically different, the animal can use this property to associate the correct arm with the food. If the two arms are identical, a place learning strategy could be used instead as in the Macfarlane experiment. Finally, if no cues at all are available, say if the maze is placed in a dark room, the animal could learn simply to turn in the correct direction at the choice point.

Morris (1981) has shown that rats can learn to swim towards a platform hidden in opaque water although there is no visual stimulus to approach. In this case, the animals obviously use a place strategy. Somehow various stimuli in the room are used to identify the position of the hidden platform. This is perhaps the most elegant example of place learning demonstrated so far. While many objections can be raised against the interpretation of Macfarlane's experiment, in the case of Morris water tank, the presence of place learning is beyond doubt.
Numerous experiments exist were an animal learns to perform a specific action such as turning right at the choice point in order to receive a reward. The Skinner box discussed above is a special case of this learning type. This is certainly some sort of response learning, but whether a stimulus is involved, we do not know.

In the light of these experiments and many others like them, what can we say about stimulus–response learning? All three types of learning can be observed if the task at hand makes demand on them. We have seen that something much more complex than a response is often learned and that a stimulus need not even be present at times. But does the list stop here or are there other types of learning as well?

We will not pursue this question here but simply conclude that if there is *one* general learning mechanism, it needs to be much more advanced than stimulus–response learning. Perhaps the reason that it has been so hard to find *the* learning mechanism in animals is simply that it does not exist. This leaves us with two possibilities. Either there is no learning at all or there are a number of interacting learning mechanisms.

4. INNATE MECHANISMS AND THEIR RELATION TO LEARNING

To assume that there is no learning seems absurd in light of the experiments described above. It may nevertheless be interesting to consider to what extent animals can behave without learning. Although learning has been found in almost all animals where one has looked for it, it is also well known that most behaviours do not solely depend on this ability. This is what makes a cat different from a dog and mice different from men. At this point, we must enter the area of species-specific behaviour.

Such behaviours are perhaps most well known through the use and abuse of the word *instinct*. Everything specific to a species was once called an instinct. Eventually the world was extended to explain all animal behaviour and was at the same time rendered meaningless. A more useful concept is that of *innate releasing mechanisms* as introduced by Tinbergen and Lorenz.

A classic example of such a mechanism, originally from von Uexküll, is the bite reaction of the common tick (*Ixodes rhicinus*). As described in Lorenz, (1977), "the tick will bite everything that has a temperature of +37 °C and smells of butyric acid". There is no learning involved in this behaviour. Instead, an innate releasing mechanism is used that reacts on a specific *sign stimulus* that starts a *fixed motor pattern*. Perhaps this type of innate releasing mechanisms can be used to explain almost all animal behaviour. Perhaps what we believe to be intelligence is only an amalgamation of such fixed behaviours. Can it be that learning only plays the role of adjusting these fixed behaviours to minor changes in the environment or body of the animal? Is learning

simply a process of parameter setting in an essentially fixed cognitive system?

There is a strong tradition within linguistics that consider the acquisition of grammar as an instance of parameter setting of the above type. Though I do not personally subscribe to this view in the context of language acquisition, this could certainly be the case in many other situations. Most fixed motor patterns would obviously profit from some degree of adaptation. This, of course, would no longer make them fixed.

A system of this kind that has been much studied in recent years is the *vestibulo-ocular reflex* (VOR) of many animals (Ito 1982). The role of this reflex is to keep the image on the retina steady when the animal moves. The reflex system is controlled by an essentially fixed system that monitors the position and acceleration of the head and flow of the retinal image and tries to compensate for it by moving the eyes. While the behaviour is entirely fixed, its high demands on the control circuits involved makes learning necessary. This is an example of an essentially fixed motor pattern that is constantly fine tuned. We may call a system of this kind a *parametrized motor pattern*.

Another example can be found in the 'imitation' behaviour of newborn childs. Almost immediately after birth, a child will imitate a number of facial gestures such as sticking the tongue out or opening the mouth (Melzoff and Moore 1977). While this phenomenon is often referred to as a very early ability to transform a visual cue to motor control it may as well be governed by something very similar to a sign stimulus. In either case, this ability develops over the years into something much more complex and is thus another example of an innate ability that show some degree of adaptation.

A related mechanism is the smiling 'reflex' that also can be shown in neonates (Johnson & Morton 1991). A newborn child smiles towards any visual pattern that shows some critical similarities with a human face. As the child grows older, the patterns that elicit this reaction will gradually change and will need to be more and more similar to real faces. Again, we have a behaviour that is innate but changes as a result of experience.

This phenomenon is similar in many respects to *imprinting* in animals. The animal has some innate conception of what will constitute an appropriate stimulus for the reaction but this innate template is enhanced by learning. In the case of imprinting and the well known following behaviour of, for example, geese, the learning process is very fast. The first moving object that the goose sees will be imprinted and thereafter constantly followed.

In other cases, for instance in song learning, the process is much slower and requires considerable practice (Marler 1970). The bird has an innate template that describes the approximate song of its species but the precise song must be learned from listening to other birds. If a bird is reared in an environment where it cannot hear the song of its own species, it will instead imitate the song most similar its template. If it does not hear any song sufficiently similar to this template, singing will not develop much.

There are currently two influential ideas that are of great importance for the relation between innate abilities and learning. The first is the concept of *preparedness* introduced by Seligman (1970) and the second is the existence of *species-specific defence mechanisms* as proposed by Bolles (1970).

Seligman challenges what he calls the *assumption of equivalence of associability*. This is precisely the assumption that was the driving force behind the behaviourist tradition. It has turned out, however, that some associations are easier to learn than others. (See Seligman 1970, and Garcia and Koelling 1966, for examples.) Seligman suggests that we may understand associability in terms of a dimension of *preparedness* (Figure 2). An animal is said to be prepared for associations that are easily formed while it is *contraprepared* for associations that are hard or impossible to learn, that is, it is prepared *not* to learn the association. In the arbitrary experiments of the behaviouristic tradition, the animal is typically unprepared for the task. Ethologists, on the other hand, do typically study situations in nature were the animals are well prepared. This can make the difference between perfect learning in one trial and no learning in 1000 trials.
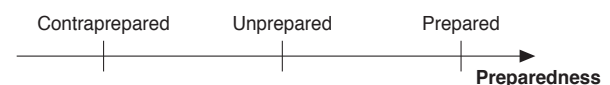

Figure 2. The Preparedness Dimension

A classical example of preparedness was demonstrated in an experiment by Garcia and Koelling (1966). Rats were allowed to drink 'bright, noisy water' and later confronted with its dreadful consequences. The water was made bright and noisy by a device that would flash a light and make a noise as soon as the animal came into contact with the water. After drinking this water, one group of rats was given electric shock. Another group was instead made ill by being injected with a toxic substance. Two other groups of rats were allowed to drink saccharine tasting water. One of these groups was also given electric shock while the other was made ill.

While testing the animals the next day it was observed that the rats that had been drinking bright, noisy water and later received shock had learned an aversion for the water. On the other hand, the group that had been made ill did not show any aversion for the water. Obviously, rats do not consider a flashing light or a loud noise a cause of illness. This result was elegantly balanced by the other two groups. Of these, the group that had been made ill showed an aversion for saccharine tasting water while the other group was unaffected. Thus, taste is easily associated with illness and lights and noises are easily associated with shock. Associations between light and illness or taste and shock are however very hard to acquire (figure 3).

Noise + Light $\xrightarrow{\text{Easy}}$ Fear

Noise + Light $\xrightarrow{\text{Hard}}$ Illness

Taste $\xrightarrow{\text{Hard}}$ Fear

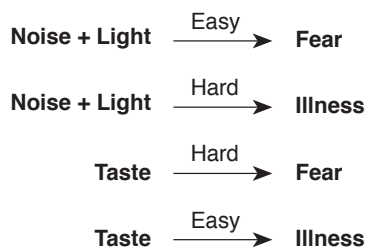Taste $\xrightarrow{\text{Easy}}$ Illness

Figure 3. The potential associations in the Garcia and Koelling experiment.

It has been pointed out that the equivalence of associability is not required by the traditional approach (Timberlake 1984). It was this assumption, however, that led the researchers of the time to study rats and pigeons in order to learn more about human learning and while it does not require the equivalence of associability it does not offer any explanation for the differences either. There is also an unusual systematicity in the associability that is out of reach for the traditional approach.

For example, it is very hard, and in many cases impossible, for a rat to learn to press a bar to avoid shock. Other behaviours such as running are learned almost immediately. In an influential paper on the subject, Bolles (1970) suggested that just as animals have specific behaviours for other engagements such as eating, obtaining food and mating, they must also have innate defence behaviours.

> "Such behaviours must be innately organized because nature provides little opportunity for animals to learn to avoid predators and other natural hazards. A small defenceless animal like the rat cannot afford to learn to avoid these hazards; it must have innate defence behaviours that keep it out of trouble" (Bolles 1978, p. 184).

The hypothesis is that associations that are in agreement with the species-specific defence mechanisms (SSDMs) are easily learned while others are much harder or even impossible to acquire. To receive food, a pigeon will easily learn to peck at a bar since pecking is in agreement with its innate eating behaviour and consequently in agreement with food. But this behaviour is highly incompatible with its innate avoidance mechanism and will thus only with great difficulty be associated with shock evasion. We see that here we have a possible explanation for the variability of preparedness as suggested by Seligman.

There are even cases where the SSDMs may hinder the animal from performing the response to be learned. This is the case, for instance, when the frightened rat freezes instead of pressing the lever in the Skinner box. Another striking example of the role of SSDMs have been shown in a modified version of the experiment where a rat has to avoid shock by pressing a bar. In this experiment, pressing the bar would remove the rat from the box and would consequently let it avoid the shock. In this variant of the experiment, the rat could easily learn to press the bar (Masterson 1970). Getting away from the box could apparently reinforce bar pressing while simply avoiding the shock could not. Considering these examples it is hard to understand how the behaviourists where ever able to teach their animals any of their arbitrary behaviours.

> "The truth of the matter is that our finest learning researchers have been keen observers of the organization underlying an animal's behaviour; they simply incorporated their observations and knowledge into the design of their apparatus and procedures rather than into their theories. It is this talent in observation, as much as the power of the accompanying theoretical analyses, that has made the arbitrary approach so viable. A truly arbitrary approach to animal learning would have failed long ago, as it has for countless pet owners, parents, and students in the introductory psychology laboratory" (Timberlake 1984, p. 183).

We may conclude that there exist a large number of innate behaviours that interact with learning in a highly complex way. These innate behaviours may either make learning easier or harder. There also exist innate preferences for forming some associations and not others. Again we see that there is nothing general about learning. The supposedly general law the behaviourists tried to discover was the result of the arbitrariness of their experiments. For an arbitrary experiment the animal is generally unprepared and can be supposed to learn slowly and regularly. In nature, however, the animal is well prepared for the types of learning that it will be confronted with. The mechanisms involved in these situations may be entirely different.

5. Interacting learning systems

In a recent learning experiment, Eichenbaum *et al.* (1991), have shown that rats will learn to categorize odours without being reinforced for doing so. Rats that were trained to discriminate between odours on a first trial were no more successful at a second trail than rats that had initially been exposed to the same odours without reinforcement. On the other hand, both these groups performed better at the second trial than did rats that had not been previously exposed to the odours at all.

A conclusion that can be drawn from this experiment is that there exist two distinct learning mechanisms that are used in the discrimination task. The first mechanism is concerned with the categorization of odours while the second mechanism is used to associate odour categories with the appropriate responses. Learning by the second system is typically performed on a single trial once the odours are known while the first system is somewhat slower. This would explain why prior exposure to the odours speeds up learning regardless of whether or not discrimination is reinforced. What we have here is an example of perceptual categorization as a process independent of response learning.

It should be noted that there exists some evidence that at first may seem to be in conflict with this discovery. Skarda and Freeman (1987) report changes in the EEG of the olfactory bulb as a result of reinforcement. Since the bulb is generally assumed to be responsible for olfactory categorization, this finding seems to indicate that the categorization process is influenced by reinforcement. Such a conclusion rests, however, on the assumption that physical areas of the brain can be identified with specific learning systems and this needs not necessarily be correct.

The idea that there exists more than one learning system is not new. Even among the behaviourists, there were researchers that belonged to this position. Clark Hull, for example, postulated (at times) that two interacting learning systems were needed to explain the experimental data. In the *primary* system, learning was induced by reduction of drive, while the *secondary* system was controlled by conditioned reinforcers, that is, events that had acquired reinforcing properties through classical conditioning (Hull 1952).
While Hull's two systems are no longer considered an accurate model of learning, they do show that not all behaviourists believed in one general learning system. It should be noted that Hull was one of the few early psychologists that wes more interested in

fitting the theory to data than selecting data that supported the theory.

> "Hull's willingness to be wrong was a remarkable, perhaps unique, virtue. It is a virtue that is, unfortunately, not shared by many theorists" (Bolles 1978, p. 104).

6. The role of reinforcement

We have seen above that learning of odours can occur entirely without reinforcement although this learning may not be expressed in behaviour until reinforcement is introduced. During the 1950s, the role of reinforcement was one of the most intense research areas within learning theory. Hull had made the entirely sensible, but as we now know, insufficient, assumption that an animal will learn to perform an action if its internal drive or need is reduced. For example, a hungry rat that is allowed to eat after having pressed a bar will reduce its hunger drive. Drive-reduction would then reinforce bar pressing. This *drive-reduction hypothesis* became one of the most influential ideas in psychology ever.

In one of Tolman's most famous experiments (Tolman & Honzik 1930), a number of rats were allowed to run in a maze for several days. One group was rewarded at the goal box while one group did not receive any reward. After the 11th day and thereafter, both groups were given food reward in the goal box. At this point the, previously unrewarded rats began to perform as well as the rats that had received reward all along. The unrewarded rats had obviously learned as much about the maze as the rewarded rats. But the learning was not expressed until reinforcement was introduced. This phenomenon is known as *latent learning*.

Figure 4 exemplifies the learning curves in a idealized latent learning experiment. Group A is rewarded from the beginning and group B and C are rewarded at a later time. The performance of group A increases steadily but the performance of group B and C jumps rapidly towards that of group A when reward is introduced. Since the performance of groups B and C almost directly approaches that of group A, the learning in these groups must have been affective even before the reward was introduced. According to the reinforcement view of learning, the performance curves for group B and C should be equal to that of group A and not steeper.
There are also many situations where it is hard to define exactly what the reinforcer should be. Avoidance learning is one such case.

> "By definition, the avoidance response prevents shock from occurring, so we cannot

point to the shock as a potential source of reinforcement. On the other hand, it is not satisfactory to cite the nonoccurrence of shock as a reinforcer because, logically, there is a host of things that do not occur, and one is hard put to say why not being shocked should be relevant, whereas, say, not being stepped on is irrelevant" (Bolles 1978, p. 184).

The explanation of learning in these cases may again be caused by interaction with species-specific defence mechanisms.
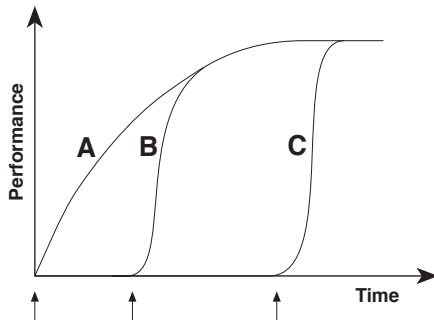

Figure 4. The idealized learning curves of three groups (A, B, C) in an latent learning experiment. The small arrows show when reward is introduced for each of the groups. When this happens, the performance of each group approaches that of the group that has received reward all along.

An alternative to the drive-reduction hypothesis is that it is the occurrence of certain stimuli that are reinforcing. This was the mechanism behind reinforcement in Hull's secondary learning system. Could all learning be explained by this mechanism? If an animal can respond to a number of innately reinforcing stimuli, then perhaps all learning could be derived from the effect of these reinforcing stimuli.

Contrary to the idea that only *stimuli* have reinforcing properties, Premack (1971) has proposed that all experiences have different values that can be used as reinforcement. The value of an activity is proportional to the probability that an animal will engage in that activity. The *Premack principle* states that access to any more probable activity will reinforce any less probable activity.

This principle was tested in an experiment where children were allowed either to eat candy or play with a pinball machine (Premack 1965). In the first phase of the experiment, it was recorded how long the children engaged in each of these activities. In the second phase, access to one activity was used as reward for performing the other. It turned out, as the Premack principle would imply, that the children that were initially more likely to eat candy than to play pinball would play pinball in order to be allowed to eat candy. The other children were, however, unaffected by the candy. Thus, candy had

only a reinforcing effect when it was used to reinforce a less probable activity.

This view of reinforcement is very different from the traditional view of Thorndike and Hull. While possibly more general, it is very hard to see how this principle can be explained in mechanistic terms. There also exists a number of cases were the principle does not hold (see Dunham 1977). It appears that reinforcement does play a role in some but not all learning. The exact nature of reinforcement is however still in debate and will probably continue to be so for a long time.

7. WHAT DOES THE ANIMAL LEARN?

What is learned when an animal in a maze succeeds in running the shortest path from the start to the goal box? Has it learned to perform a fixed sequence or motor patterns or has it constructed a cognitive map of the maze? Perhaps it has learned to expect food at a certain place or to expect reward for running a certain route. The theories are almost as many as the researchers in the field. However, there are some main directions that I will try to summarize in this section. Here I will only consider what is learned and not how that learning has come about.

STIMULUS–RESPONSE ASSOCIATIONS

The most trivial explanation is that the animal has learned a stimulus–response association. Each place in the maze is considered to give rise to a specific stimulus that is associated with the correct response to perform at that place. A problem with this approach is that the behaviour generated is unstable. The actions performed are defined as movement away from stimuli and not towards stimuli, but this is not a uniquely defined direction. The response R_0 performed as a result of observing stimulus S_0 may give rise to different movements in space depending on the initial position of the animal. Thus, S–R behaviours are *divergent*. As a sequence of S–R associations is performed, the error will accumulate until it drives the animal off course. (See figure 3). A larger set of S–R associations makes the behaviour more stable but can never overcome the inherent instability of this type of learning. It should be noted, however, that few researchers nowadays refer to this type of simple motor-pattern when they talk about responses.
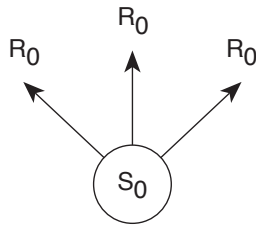
Figure 5. The result of performing a response, R_0, depends on the initial position of the animal.

STIMULUS-APPROACH ASSOCIATIONS

Another explanation may be that the animal has learned to approach a number of stimuli in the maze (figure 6). To get to the goal it has to first approach stimulus, S_0, then stimulus, S_1, and so on until it is able to reach the goal box. Contrary to stimulus–response behaviour, stimulus–approach behaviour is stable. This depends on the fact that an approach behaviour consists of a whole set of responses, A_0, which all drive the animal nearer to the stimulus. An error in the initial position of the animal will decrease as it approaches the stimuli (Figure 4). As a consequence, stimulus–approach behaviour is *convergent*. This makes this type of learning much more likely as a basis for adaptive behaviour.

This constitutes the first of the three mechanisms discussed above in relation to the simple T-maze. Stimulus–approach associations could be used to guide the animal if the two arms of the maze looked different or could be distinguished in any other way.

These structures should not be confused with what Hull (1934) called habit-family hierarchies although they are similar in some respects. A habit-family hierarchy is a set of responses or chains of responses that have the same starting point and the same goal response. Stimulus–approach structures are only concerned with goal stimuli and cannot be divided into a discrete set of responses.
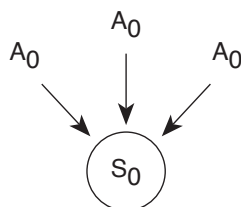

Figure 6. The stimulus S_0 can be approached from many directions, A_0.

PLACE-APPROACH ASSOCIATIONS

Place approach is similar to stimulus–approach in that it is stable but instead of approaching a stimulus, the animal uses a set of stimuli to identify a place to approach. This type of learning is more advanced than the other since it requires the ability to use a configuration of stimuli to identify a place – a far from trivial task. There exist a number of models of this process. (See for example Schmajuk and Blair 1993 and Zipser 1985.) Figure 7 shows a place-approach situation. This is the second of the possible mechanisms used in the T-maze discussed above.
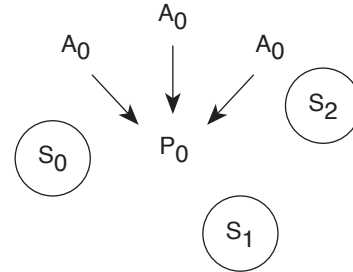

Figure 7. A place, P_0, defined by a set of stimuli (or landmarks), S_0, S_1, S_2, can be approached from a number of directions, A_0.

RESPONSE CHAINS

According to this position, what the animal learns is simply to perform a sequence of responses, $R_0, R_1, ..., R_n$, in order to move from the start to the goal. The only stimulus involved is the one that starts the chain (Figure 8). Obviously, this type of behaviour is even more unstable than a simple S–R reaction. It is very unlikely that a structure like this is used if the animal can choose other strategies.

There are nevertheless a number of situations where response chains are the only possibility. This is the case, for instance, when a fast arpeggio is played on the piano. Each new key on the piano must be pressed before any feedback signal from the fingers have had the time to reach the brain (Carpenter 1984). This means, of course, that simple stimulus–response associations must also exist as a limiting case of response chains. We have here the third of the already discussed possible mechanisms used in the T-maze.
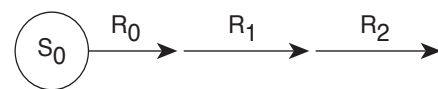

Figure 8. A response chain. A single stimulus, S_0, triggers a whole sequence of responses: R_0, R_1, R_2.

STIMULUS-APPROACH CHAINS

Just as responses can be linked together in chains, it is also possible for approach behaviours to be linked (figure 9). Like a simple stimulus–approach behaviour, these chains produce stable behaviour but they can range over much larger distances than a simple stimulus–approach association.
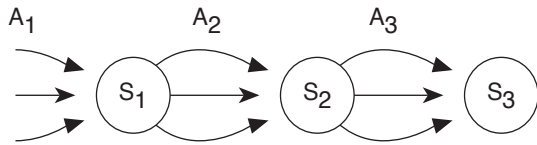
Figure 9. An approach chain. A sequence of stimuli, S_0, S_1, S_2, are approached in order.

PLACE-APPROACH CHAINS

Naturally, place-approach associations can also be linked in chains (Figure 10). Using this type of structure, the same stimuli can be used many times to locate different places. In the figure only three stimuli (or landmarks) are used to locate and approach all three places p_1, p_2, and p_3.
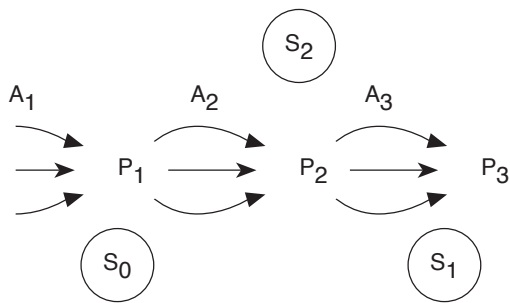

Figure 10. A place chain. A sequence of places (P_1, P_2, P_3) can be approached by identifying the relation between various stimuli $(S_0, S_1$ and $S_2)$.

S-R-S' ASSOCIATIONS

The types of associations described above can be used to control behaviour but they cannot be used to make inferences. Tolman, postulated that animals learn something like S–R–S' associations (*cf.* Tolman 1932). These tell the animal that if it is in situation S and performs response R it will end up in situation S'. Such associations are much more powerful than the others we have so far considered. For example, if the animal is in possession of the two associations S_0–R_0–S_1 and S_1–R_1–S_2, it can, at least potentially, infer that by performing the responses R_0 and R_1 at S_0 it will reach S_2. Thus, it can perform sequences of responses in order to obtain a goal even if that particular sequence has never been performed before.

By acquiring sufficiently many S–R–S' associations, it is possible to build a topological map of the environment (Figure 11). This map can be used with great utility in shortcut and detour problems as well as in general problemsolving. It can also be used to detect when a response does not result in the expected situation. This is in other words a form of expectation learning.
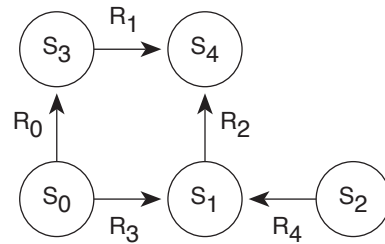

Figure 11. A map consisting of S–R–S' associations that predict the consequences of performing a responce in a certain situation.

These types of structures can be further extended by assuming that the animal has the ability to reverse the direction of a S–R–S' association. In this case, every time the animal knows that it can transform situation S to situation S' by performing response R, it also knows that it can transform situation S' to situation S by performing R^I, where R^I is the inverse of R.

An especially important class of systems can be constructed if we embed stimulus–approach structures within S–R–S' associations. The behaviours generated by a system of this kind are stable while at the same time supporting various forms of inferences. Systems of this kind have been proposed by Gallistel (1980) and also by Schmajuk and Thieme (1992).

S-R-S* ASSOCIATIONS

Another possibility is that the animal learns to associate a stimulus, S, followed by a response, R, with a certain motivationally significant stimulus S*. If S* is a stimulus that gets more intense as the animal approaches a goal, associations of this type could be used to guide the choice of responses at S. The response associated with the most intensive S* should be selected in order to reach the goal.

Like, S–R–S' learning, this is a type of expectation learning, but here it is an expectation of reward and not an expectation of a subsequent stimulus that is learnt. As we will see below in section 16, a combination of these two types of expectation learning can be very powerful.

S-S' LEARNING

We will finally consider associations between stimuli. In classical conditioning, it has sometimes been assumed that it is not an association between stimulus and response that is formed but rather an association between the two stimuli involved. In this view, Pavlov's dog does not salivate because the bell has been associated with salivation but rather because the bell has been associated with food which in turn activates salivation. This is called the *stimulus-*

substitution theory of conditioning (Mackintosh 1974).

There are a number of processes that have S–S' associations as their basis. In *categorization*, a stimulus representing an instance of a category is associated with a stimulus representing its category. When the stimulus is perceived its corresponding category is activated. Of course, stimuli are here considered as something internal to the organism and not as external cues. We are, in fact, talking about *representations* of stimuli. This view of learning is similar to the early associationistic school that considered associations as links among ideas. Hebb's cell assembly theory is a more sophisticated variation on this theme (Hebb 1949).

The above list is by no means exhaustive. I have only touched on some of the most important ideas about what is learned by an animal. Numerous attempts have been made to explain each of the above learning types by means of the other but so far there is no consensus in the area. The view I am advocating is that all these learning types, and perhaps many more, co-exist and interact with each other during learning and behaviour.

8. INTERNAL INFLUENCES ON BEHAVIOUR

So far, I have described behaviour as if it were guided primarily by external stimuli. This is of course not the case. Internal determinants of behaviour are very prominent in most situations.

One obvious internal determinant is the current need of an animal. In identical external situations, a hungry animal will eat if possible while a satiated animal will not. Internal stimuli are related to the concept of *motivation*, but since this determinant of behaviour is not directly relevant for the present argument, I will not dwell on this matter here. I have so far assumed that there is only one goal to pursue and that the animal is motivated to do so.

Another determinant that is more relevant to the present argument is what I will call the *internal context* of a situation. In many learning paradigms, the appropriate action for a given situation depends on some previous action performed at the same place or in the same situation. To make the correct choice of an action at the second trial, the animal must remember what it did the last time. The internal context of a situation is the internal state that somehow reflects this previous choice.

In Olton's radial maze, a rat is supposed to visit each arm of a maze once and to learn this behaviour, the rat receives a reward on its first visit to each arm (Olton and Samuelson 1976). Each time the rat is in the centre of the maze, it has to choose a new arm to visit (figure 12). Since the rat cannot perceive the reward from the centre of the maze, this behaviour seems to require some memory for the previously made choices.

Rats are surprisingly good at this task and remember which arms they have visited without much trouble. This is the case even in very large mazes with sometimes as many as eighteen arms. They do not, however, follow an obvious strategy like selecting each arm sequentially around the maze but move around seemingly at random. It is interesting to note that the demands on memory required for this solution is clearly out of reach for most humans.
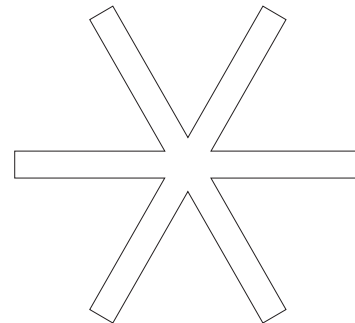

Figure 12. A radial maze with six identical arms. The animal is rewarded on the first visit to each arm. To be successful, the animal must remember what arms it has already visited.

As a determinant of behaviour, the internal context is no different from external stimuli. It is used to direct behaviour in exactly the same way but it differs in the way it is generated. External stimuli are gained through the perceptual apparatus of the animal but the internal context has to be generated from other sources. One possible mechanism is a *working memory* that stores the actions previously performed by the animal (Olton and Samuelson 1976).

While it is clear that some sort of memory is necessary for these types of tasks, it is not at all established what properties such a memory system must have. For instance, how is the relevant internal stimuli recollected from all the potential memories that could be relevant in a given situation? How does the animal decide on what to store in memory? Whatever properties a learning system involved in this type of memory may have, it must interact with the different learning strategies I have presented above.

9. ARE THE INTERNAL STRUCTURES AN IMAGE OF REALITY?

Assuming that an animal behaves in an appropriate way, does this mean that it knows something about

its world? It is tempting to assume that a rat that has learned to run through a maze to receive food does so because it is hungry but would prefer not to be. It knows where the food is located and how to get there and expects to be less hungry if it eats the food. Based on this information, the rat can infer that the best way to satisfy its goal is to run through the maze and eat the food, and as a consequence of this inference it will decide to run through the maze and eat the food.

According to Tolman (1932), this is an adequate description of what goes on in the mind of the rat and it is not hard to understand Guthrie's objection that according to this view the rat would be "buried in thought". However, the main criticism of this view has not come from within animal learning theory but instead from ethology and ecological psychology.

When the smell of butyric acid with a certain temperature causes the tick to bite, there is no reason to believe that it has some objective knowledge of mammals that is used to decide on whether to bite or not (Sjölander 1993). In fact, it seems inappropriate to talk about knowledge at all in this context. In nature, everything that smells of butyric acid and has a temperature of +37 °C is a mammal and in the world of the common tick, this is all that a mammal is.

The part of reality that is within reach of the perceptual apparatus of an animal can be referred to by the concept of *Umwelt*[3] as proposed by von Uexküll. There is no reason to assume that an animal has a better conception of reality than is necessary. The *Umwelt* of the common tick is not very sophisticated but is sufficient for it to survive. If the tick believes that everything that smells of butyric acid is something it should bite, it will survive, if it does not it will probably die. This does not mean that its conception of reality is true in any objective sense but this is not terribly important as long as it significantly increases the chance of survival for the animal. It is sufficient for the concepts of an animal to *parallel* the real world. They do not have to be isomorphic to it (Sjölander 1993).

In ecological optics (Gibson 1979), the idea of an *ambient optic array* is used in a way that is very similar to an *Umwelt*, but while this concept refers to all aspects of the environment, the ambient optic array refers only to the visual surrounding of an animal.

Ecological psychology emphasizes the role of *invariants* in the environment that can be directly picked up by an organism. The sign stimulus that causes the bite reaction in the tick is an example of such an invariant. As pointed out by Runesson (1989),

---

[3]Umwelt means approximately surrounding environment.

it is sufficient that invariants are incomplete, that is, they should hold sufficiently often for the mechanisms that rely on them to be adaptive. This is certainly the case with the sign stimulus of the bite reaction.

In the behaviourist accounts for learning it was often implicitly assumed that animals perceive the same (objective) world as humans. No-one was ever surprised to find that animals attended to exactly those stimuli that were relevant for the learning task. For some reason, the world of the animals coincided with that of the experimental situation. As a consequence, only those stimuli specially prepared for the learning task needed to be considered when attempting to explain learning.

In the light of the example above, this should be very surprising. Why should a rat care about exactly those stimuli that were needed to solve the problem and not on something entirely irrelevant like the smell of the experimenter? Of the classical learning theorists, only Pavlov considered this problem in any detail (Pavlov 1927).

None of the different learning strategies presented above gives rise to objective knowledge of the world. Some of the learned structures even depend on the learning animal in some unusual ways. For example S–R–S' association are based on the behavioural repertoire of the animal. It will not learn that A is north of B but rather that some specific action is appropriate for moving from A to B. But a structure of this kind is much more useful than an objective representation if the animal wants to move from one place to another.

10. WHEN DOES THE ANIMAL LEARN?

In this section we will consider how different proposed learning mechanisms relate to the obtainment of a reward. Learning has been described as occurring either before, after or at the same time as a reward is received. I will call these different learning types *early*, *synchronous* and *late* learning (See figure 13).
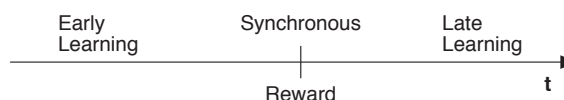
Figure 13. Different times when learning can occur.

EARLY LEARNING

Early learning is learning that occurs prior to the obtainment of a reward. If a rat learns the location of food without being allowed to eat it, we have an

instance of early learning. Thus, early learning is involved in latent learning experiments. We may hypothesize one of two distinct processes responsible for early learning.

The first process, usually associated with Tolman, explains learning simply as the gathering of information about the environment. The construction of a cognitive map is an example of such a process. Both S–R–S' and S–S' associations can be constructed using this type of early learning. It is important to note that the demands on the cognitive apparatus that an animal needs for this mechanism are rather high. Consequently, we would only expect to find this type of learning in higher animals.

The second process is driven by the distance to a goal object. An anticipatory reinforcement signal is generated that is inversely proportional to the perceived distance to the goal object. The closer to the object, the larger the reinforcement will be. In this case, an animal will learn to approach food even if it is not allowed to eat it. A learning mechanism of this type implies that maximal reinforcement will be received when the goal object is actually reached.

While this type of learning has many important merits it critically depends on a competent evaluation of the distance to the goal. Perhaps it is the failure to perceive this distance that makes the dedicated gambler risk even more money after 'almost winning the bet'. As far as I know, this type of learning has not been studied in the animal learning literature.

Since early learning does not depend on any reward, phenomena like latent learning are easily explained with either of these learning mechanism. In the case of shortcut and detour behaviours, it seems that the first learning mechanism is necessary.

SYNCHRONOUS LEARNING

Synchronous learning is perhaps the most obvious alternative to the drive-reduction hypothesis. Here it is the consummatory response that is the origin of learning. When an animal eats the food, its previous responses are reinforced. Among the classical learning theorists, Guthrie is the main proponent of this view.

It does not appear that synchronous learning can explain the more complex behaviours of an animal but there are some situations where a mechanism of this type seems most appropriate. For instance, learning the correlation between smell and taste is obviously best done when both types of information are present, and this is only the case while eating.

LATE LEARNING

Hull's drive-reduction hypothesis is a classical example of late learning. Here it is not the reward itself, such as the food that causes learning, but rather its consequences on the organism. According to this hypothesis, the reduction of hunger would reinforce learning while eating should not.

How are we to choose between these learning types? Again, I want to propose that they are all effective but in different circumstances. In many cases, early learning is certainly the case, but can that type of learning explain all cases where behaviour is changed? Because of the complexity involved in early learning it is not entirely unrealistic to assume that there also exist less complex learning mechanisms such as synchronous and late learning. At least in simpler organisms, these are the mechanisms to look for.

We may also make the conjecture that if these less sophisticated learning types are present in simpler organisms, they are also very likely to play a role in more advanced organisms. After all, they are still entirely sensible.

11. SUMMARY OF ANIMAL LEARNING

I hope to have shown that learning in animals is a highly complex and complicated business. It is quite unlikely that all the examples described above can be explained by one mechanism and if it can, it is certainly very different from any of the currently proposed learning theories.

In summary, there are a number of important facts about animal learning that we must consider if we want to construct or model an intelligent system.

◊ It is unlikely that there exists one general learning mechanism that can handle all situations. Animals are prepared to learn some associations and not others.

◊ Many alternative strategies are available to use for the same problem, like place learning, approach learning or response learning. The strategies are selected and combined according to the demands of the task.

◊ Animals have a number of species-specific mechanisms that interfere with learning. Such innate behaviours are necessary to keep the animal alive while it learns about the world. In some animals, almost all behaviours are of this kind.

◊ What the animal learns can be represented in a number of ways. We have seen at least nine ways to represent habit and knowledge. These structures need not be good descriptions of the external world. It is sufficient that they help the animal stay alive.

◊ Memories of past actions or experiences are sometimes necessary to choose the correct behaviour. These memories are probably different from the other structures learned by an animal.

◊ Learning can occur at different times with respect to a reward. Learning that occurs prior to any reward is in effect independent of the reward but can usually only be demonstrated once a reward is introduced.

12. PARALLELS BETWEEN ARTIFICIAL INTELLIGENCE AND ANIMAL LEARNING THEORY

It is interesting to see that many artificial intelligence models show striking similarities with the animal theories. The reinforcement theories proposed by Thorndike and Hull find their counterpart in the early learning algorithms such as the one used in Samuel's checkers program (Samuel 1959) and more contemporary reinforcement learning models (Barto and Sutton 1990). The parallel of Tolmans' theory can be found in mainstream artificial intelligence in the use of internal world models and planning. We also find the equivalent of the ethological approach to animal behaviour in the work of Brooks and others that emphasize the role of essentially fixed behavioural repertoires that are well adapted to the environment (Brooks 1986).

These similarities have made me curious to see whether it would be possible to match the different fields and perhaps transfer ideas between them. Can insights from animal research be used to construct intelligent machines? Is it possible that research on artificial intelligence has anything to say about how animals and humans work? I think the answers to both these questions are affirmative and the present work is partly an attempt to carry out such a matching.

In the following sections, we will take a closer look at the different learning methods used by various artificial intelligence researchers and try to match them with the relevant animal learning theories. The result of this exercise will be an attempt to formulate some general design principles for an intelligent system.

13. S–R ASSOCIATIONS IN AI AND CONTROL

The rules used in rule based systems are very often similar to S–R associations. When one rule is used to generate the precondition for another rule, the process is not entirely unlike the chaining of S–R associations. In the animal earning theories, the environment holds the result of a response and may in turn trigger the next S–R association. In rule based systems, the environment is replaced by an internal representation of 'facts' generated by the triggered rules. Computationally, the two approaches are almost identical although the languages used to describe them are entirely different.

Perhaps a clearer example of S–R associations can be found in the use of look-up tables (LUT) in both AI and control. Look-up tables are used to store the output for a set of inputs. This has the advantage that no calculations have to be made. For a given input, the result is simply looked up in the table. A control strategy can be coded once and for all in a look-up table to make the control faster than if the controlling signal had to be calculated for each input.

Look-up tables have two disadvantages however. The first is that there may exist inputs that are not stored in the table. These inputs have no defined output. The second problem has already been mentioned in relation to S–R learning: behaviour generated by S–R associations is divergent. Both these problems have been addressed by generalizing look-up tables. These data structures will interpolate between the entries in the table to find an output for an unknown input.

Albus' CMAC was one of the first mechanisms to use this idea (Albus 1975). The model was supposed to describe learning in the cerebellum and since its introduction it has been developed in two quite distinct directions. The first is in the field of control where it is the basis for many control strategies based on generalizing look-up tables (*e. g.* Atkeson and Reinkensmeyer 1990, Kraft and Campagna 1990). The other development of the model has been towards a more realistic model of cerebellar learning. Most contemporary neurophysiological models of classical conditioning have the CMAC model as their starting point (*e. g.* Ito 1989, Moore *et al*. 1989). Another connection between animal learning theory and control theory if the Recorla–Wagner model of classical conditioning (Rescorla and Wagner 1972). This model is mathematically identical to the Widrow–Hoff learning rule for adaptive filtering (Widrow and Hoff 1960/1980).

14. Reinforcement learning

Thorndike's *law of effect* states that the learning of a response is governed by the effects of that response. The cat will learn to press a lever to escape from its box since the effect of lever pressing, that is the escape, is pleasant. The pleasant aspect of escape reinforces the behaviour that precedes it. As a consequence, this behaviour is more likely to be elicited again. If, on the other hand, a behaviour is followed by some unpleasant event, the likelihood of the behaviour is reduced instead. The closer in time a response is to the reward, the more the response will be reinforced. While this description comes from animal learning theory, it is essentially the idea behind reinforcement learning as it is used in artificial intelligence.

Something similar to reinforcement learning was first used in Samuels' checkers program that was developed in the late fifties (Samuel 1959). When the computer wins a game, it receives a reward in the form of a positive evaluation of its last few moves. During later games, this evaluation is propagated towards earlier positions of the game. Moves that lead to favourable positions receive a higher reward (that is a better evaluation) than moves that are less successful. Eventually all moves will have been evaluated and the computer will be able to play the game fairly well.

While this learning scheme is feasible in principle, it will take an almost infinite amount of time before all moves have been tested. This problem was overcome in two ways. The first was to let the program use a static evaluation function on moves that were far from any known position. The second solution was to let the program use a high-level description of the positions. Using this complex description, evaluations of one position could be generalised to a position that had never been encountered before. The high-level descriptions were also further enhanced by the introduction of learning.

This idea has later been included as a component in many learning systems. The bucket brigade algorithm used in Holland's classifier systems is another instance of this general learning scheme (Holland *et al*. 1986). The learning system receives a reward from the environment and its task is to adapt its internal rule base in such a way that it receives an optimal reward from the environment.

Q-learning as proposed by Watkins (1992) is perhaps the easiest reinforcement learning algorithm to understand. The main element of this algorithm is the Q-function that assigns an expected reward to each combination of a situation (or stimulus) and an action (or response). When the system finds itself in a certain situation, it simply chooses the action for which its expected reward is largest. In effect, the Q-function describes a set of S–R–S* associations. The role of the learning algorithm is to construct an estimation of the Q-function by trying out the different actions in the environment.

Common to all of the above examples of reinforcement learning is that actions that are not immediately rewarded are reinforced by the actions that follow them. The propagation of reward from the terminal action towards the preceding ones is not entirely unlike the anticipatory goal reaction, r_G, proposed by Hull (1943, 1952). This reaction, whose only effect would be to generate an anticipatory goal stimulus, s_G, would initially be associated with the rewarding response and would later propagate through the chain of S–R associations and serve as the glue in a response sequence.

The connection between animal learning theory and reinforcement learning has recently been emphasized in a number of articles by Barto and Sutton (Barto, Sutton and Watkins 1990, Sutton and Barto 1990). Their temporal difference method has been used both as a biological model and as an adaptive control strategy and is one of the most recent attempts to make the connection between animal learning and control theory explicit. Baird and Klopf (1993) describe a modified version of the Q-learning paradigm that also clarifies this connection. They show how Q-learning can be changed to conform with the precise details of several animal learning experiments.

15. Learning as knowledge acquisition

According to Tolman, learning is the acquisition of knowledge about the world. This view is the most popular among most contemporary psychologists and AI researchers, and there exists an endless number of models and systems based on this approach. Reasoning and problem solving are examples of abilities that seems to require knowledge. Based on knowledge of the world, we are able to reason about the outcomes of actions, we can solve problems and make plans.

The solution to many spatial problems requires that the animal in the maze makes some form of inferences about what route to take from the start to the goal box. In a classical experiment of Tolman and Honzik (1930) a rat is allowed to explore the maze in shown in figure 14. After some practice the animals will use the straight alley from the start box, S, to the goal

box, G. Once this habit is formed, the path from S to G is blocked at point B in the maze. Consistent with reinforcement theory, the rats now chose the next shortest path on the right of the maze. When the direct path is instead been blocked at point A, according to reinforcement theory, the rats would now try the second shortest path on the right instead. This does not happen, however. Instead they will directly choose the longest path on the left.

This is, of course, the most sensible choice since the right path is also blocked at A but to make the correct choice some considerable cognitive abilities are required. Its seems that some kind of internal world model is required and that the animal uses this model to infer that the right path will also be blocked before it chooses the right one.
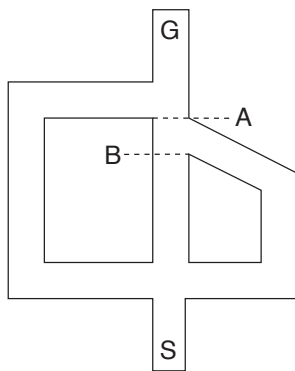

Figure 14. The maze of the type used in the Tolman and Honzik (1930) experiment.

Tolman's view that learning is essentially the acquisition of knowledge about the environment has no problems explaining this behaviour, nor do most artificial intelligence systems for planning and problem solving. If the world is represented as a set of S–R–S' associations, the choice of the correct path is given by invalidating the S–R–S' association that leads past the point A where the path is now blocked and replan the route from S to G.

Most AI planning systems make use of representations that are very similar to S–R–S' associations. They are usually of the form:

$$precondition \times action \rightarrow outcome.$$

A plan is simply a sequence of these representation where the outcome of one rule is the precondition for the next (Fikes and Nilsson 1971). Planning and problem solving is thus reduced to a search for a sequence of rules that leads from the start to the goal. If such a sequence is found, it is subsequently executed in the real world.

The planning process can be made more efficient by building new rules that describe the combined result of executing several actions in succession. If the planning system finds two rules, $x{:}a \rightarrow y$ and

$y{:}b \rightarrow z$, it can combine these into a new rule $x{:}b{\circ}a \rightarrow z$. The next time the planner wants to go from x to z no planning is necessary. This process is called *chunking* and has been much studied in the cognitive literature. (See for example Newell 1990.) As a result of chunking, the planner will become better with added experience.

The view that all behaviour can be described in this way have received much criticism in recent years and many of the deficiencies of these type of mechanisms have been acknowledged (*e. g.* Maes 1990b). For example, it is often the case that once the planning process has finished, the rules used to construct the plan may no longer be valid. There are nevertheless many situations where a problem solving ability seems to be necessary. This has lead some researchers to try to combine the reactive approach with planning in different ways. One of the greatest insights gained from this work is that plans should be considered more as resources than as programs to execute (Payton 1990). The immediate sensory readings from the environment should always have precedence over an internal plan.

There have also been some attempts to combine an internal world model with reinforcement learning. The DYNA architecture proposed by Sutton (1992) is one noticeable example of this. Using an internal world model, the agent can try out actions internally instead of confronting them with the cruel and unforgiving results of reality. While these internal tests are performed, the reinforcement learning system will adapt and the appropriate actions can then be executed externally. It has been shown that this approach speeds up Q-learning considerably (Peng and Williams 1993). This is an example of a model where S–R–S' learning, (the internal model) is combined with S–R–S* learning (the Q-function).

Another important role of planning is to anticipate future states of the world (Rosen 1985). This ability makes it possible to let anticipated future states of the world influence the present behaviour of the agent. For example, an animal that anticipates its own future needs may gather food even before it becomes hungry (*cf.* Gulz 1991).

In summary, most behaviour of an animal may be governed by rather simple mechanisms but they also have the ability to solve rather complex problems in some cases. This ability seems to require knowledge of some kind and this knowledge must be acquired by learning. There are plenty of models within AI that may be used as a starting point for models of these phenomena.

16. The reactive approach

The view that animal behaviour is best described by a number of interacting innate motor patterns has been the inspiration for the currently most fashionable approaches to robot control.

> "[T]he emphasis in these architectures is on more direct coupling of perception to action, distributedness and decentralisation, dynamic interaction with the environment and intrinsic mechanisms to cope with resource limitations and incomplete knowledge" (Maes 1990a).

The most important aspect of such architectures is their emphasis on complete creatures or systems that let us make observations that cannot be made from studies of isolated modules (Brooks 1986, 1991a, 1991b).

The subsumption architecture introduced by Brooks (1986) is a computational model that is based on a network of asynchronously computing elements in a fixed topology. The active elements communicate with each other and with sensors and effectors by sending and receiving messages without implicit semantics. The meanings of the messages are given by the operations of both the sender and the receiver (Brooks 1991b). Typically, the messages are constrained to be very small values represented in a low number of bits. The communication rate is usually very low, on the order of a few messages every second.

The robots built according to these principles differ from more traditional designs in that they are behaviour based (Connel 1990, Horswill and Brooks 1988). In this context, a behaviour is a subsystem that is responsible for some specific action pattern of the robot. There are many connections between this approach and models in ethology. For instance, the behaviours of the robots are similar to fixed action patterns.

There are also a number of similarities between the perceptual systems of these robots and the idea of direct pick up in ecological optics. For instance, Horswill (1992) presents an interesting analysis of the visual invariants in an office environment that is directly inspired by the ecological approach.

The most radical defenders of this view deny the need for any type of internal representations or reasoning mechanisms (Brooks 1991b). Even memory is considered harmful since it gives the robot an internal state. Since internal states may not adequately describe the external situation, a robot should react directly on the external world and not on some internal representation of it. This is the idea of using "the world as its own model" (Brooks 1991a).

While this may be a good idea in general, we have already seen that memory *is* necessary for situations like the radial maze. It is therefore reassuring to see that Brooks now acknowledges this need (Brooks and Stein 1993).

17. Perceptual learning

In the cognitive literature, perceptual learning is usually described in terms of concept formation and prototypicality (Rosch 1973, Glass and Holyoak 1985). Within the behaviourist school, the same phenomena are studied in the context of discrimination learning and generalisation gradients. The difference between the two views of categories can be seen in figure 15. The figure to the left shows instances of three categories with a discrimination border drawn between them and the figure to the right shows the three categories as bounded regions around the examples.
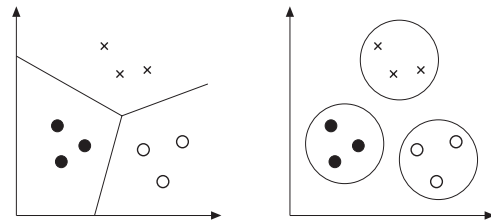

Figure 15. Two views of categories. Left. Discrimination surfaces in the instance space. Right. Categories as regions.

The main difference between the cognitive and the behavioural approaches has not been in the phenomena studied but rather in the way these phenomena have been attributed to different mechanisms. The cognitive investigators have searched for the internal representations of categories while the behaviourists have studied the tendencies to react to different stimuli. In both cases, one has found that categories cannot in general be described by sharp borders. Instead they have a radial structure where some instances of a category are better examples of that category than others.

In cognitive science, this is taken as evidence for the prototype theory. This theory states that some members of a category are more prototypical than others (Rosch 1973). For example, a prototypical chair has four legs. But there also exist chairs with three legs or perhaps only one. These are thus less prototypical, that is, less good examples of the concept of a chair.

The radial structure of categories has also been studied within the behaviouristic tradition. When the

tendency to respond to a stimulus is measured, it can usually be shown that there exists one specific stimuli for which the response is strongest or most likely (Mackintosh 1983). As the stimulus is altered the response decreases with increased dissimilarity between the optimal and the altered stimulus. There is said to be generalization gradient around the optimal stimulus. Is this not prototype theory in disguise?

Another way to study perceptual learning is to see whether an animal will react in one way or another on a stimulus. In this way we will study discrimination surfaces between different categories instead of their radial structure.

All these approaches to perceptual learning and many other can be found both within the area of machine learning (Davidsson 1994) as well as in neural networks and statistical inference (Lippman 1987).

18. Conclusion

All the studies presented above, both within animal learning theory and artificial intelligence, have been concerned with some particular aspect of learning or behaviour. To date, very few models have attempted to deal with the full complexity of learning, although there certainly exist biological models that could explain most aspects of learning if they could only be combined in some sensible manner.

Too much effort has been spent on trying to figure out who is right and who is wrong instead of attempting to see the similarities between the different theories. While the extreme positions usually get the most attention, the truth is, more often than not, somewhere in between. An attempt to merge the different theories into a coherent system would be very welcome. However, such an enterprise would have to avoid two critical traps that have caught most previous attempts.

The first trap is to believe that all learning and behaviour can be explained with a small set of principles. The result of this approach has been the construction of grand theories that set out to explain all instances of learning but are later revealed as too limited. It is not unusual for models of this kind to be both clear and elegant, but this is true only because their explanatory power has been sacrificed.

The second pitfall is to think that everyone is right and to simply combine all models one can find into one big theory of everything. This has often been the case when AI researcher have felt the need to build complete systems. The models for perceptual learning are usually highly incompatible with those for reasoning and problem solving but this has not stopped some people from combining them into so

called hybrid systems. While these systems have the advantage that they combine many mechanisms, all signs of elegance are usually far gone. Since most hybrid systems have been directed towards specific technical applications, their value as theories is also very limited.

In summary, what is needed is an approach were all the different aspects of learning can be combined in an elegant manner. I want to propose that such an endeavour must satisfy the following three criteria.

First, it must be computational. Whatever the properties are of the system we are looking for, they will be highly complex. Thus, a model that is not computational will inevitably contain many inconsistencies. Only within a computational approach are we required to specify a model in every detail, and that is absolutely necessary in this case. This implies that we must model one particular individual. There exists no general animal and a fully specified system can never general either.

Second, it must describe a complete system. A complete system includes sensors and effectors as well as everything in between. This assures that the system will be grounded, that is, all internal processes can be traced back to the peripheral systems. Like the computational approach, the complete systems approach also requires that we model one particular individual of one particular species.

Third, the system must be based on one descriptive framework, such as neural networks or a rule based system. This may not be required to build a working system, but it is a necessary feature of any attractive model. This will make it possible to describe a system in a coherent way as well as making the computational approach easier.

To conclude, I suggest that artificial intelligence learn the lessons from animal learning theory and starts to consider complete systems where a large set of interacting mechanisms are combined in a coherent manner. The study of such systems will be of great importance both for the success of artificial intelligence and for our understanding of learning and intelligence in animals and man.

Acknowledgements

References

Albus, J. S., (1975), "A new approach to manipulator control: the cerebellar model articulation controller (CMAC)", *Transactions of ASME Journal of Dynamical Systems, Measurement, and Control*, **97**, 25–61.

Atkeson, C. G. & Reinkensmeyer, D. J., (1990), "Using associative content-addressable memories to control robots", in Winston, P. H. & Shellard, S. A., *Artificial intelligence at MIT: expanding frontiers*, Cambridge, MA: MIT Press, in preparation.

Baird, L. C. & Klopf, H. A., (1993), "A hieararchical network pf provably optimal learning control systems: extensions of the associative control process (ACP) network", *Adaptive Behavior*, **1**, 321–352.

Barto, A. G., Sutton, R. S. & Watkins, C. J. C. H., (1990), "Learning and sequential decision making", in M. Gabriel and J. Moore (eds.) *Learning and computational neuroscience: foundations of adaptive networks*, 539–602, Cambridge, MA: MIT Press.

Bolles, R. C., (1970), "Species-specific defence reactions and avoidance learning", *Psychological Review*, **77**, 32–48.

Bolles, R. C., (1978), *Learning theory*, Fort Worth: Holt, Rinehart and Winston, Inc.

Brooks, R. A., (1986), "Achieving artificial intelligence through building robots", MIT A. I. Memo No. 899.

Brooks, R. A., (1991a), "New approaches to robotics", *Science*, **253**, 1227–1232.

Brooks, R. A., (1991b), "Intelligence without reason". *Proceedings of IJCAI–91*, 569–595,

Brooks, R. A. & Stein, L. A., (1993), "Building brains for bodies", MIT A. I. Memo No. 1439.

Carpenter, R. H. S., (1984), *Neurophysiology*, London: Edward Arnold Ltd.

Charniak, E. & McDermott, D., (1985), *Introduction to artificial intelligence*, Reading, MA: Addison-Wesley.

Connell, J. H., (1990), *Minimalist Mobile Robots*, Boston: Academic Press.

Davidsson, P., (1994), "Concepts and autonomous agents", LU–CS–TR: 94–124, Department of computer science, Lund University.

Dunham, P. J., (1977), "The nature of reinforcing stimuli", in W. K. Honig and J. E. R. Staddon (eds.) *Handbook of operant behavior*, Englewood Cliffs, NJ: Prentice-Hall.

Eichenbaum, H., Otto, T. A., Wible, C. G. & Piper, J. M., (1991), "Building a model of the hippocampus in olfaction and memory", in J. L. Davis and H. Eichenbaum (eds.) *Olfaction*, 167–210, Cambridge, MA: MIT Press.

Fikes, R. E. & Nilsson, N. J., (1971), "STRIPS: A new approach to the application of theorem proving to problem solving", *Artificial Intelligence*, **2**, 189–208.

Gallistel, C. R., (1980), *The organisation of action: a new synthesis*, Hillsdale, JN: Erlbaum.

Garcia, J. & Koelling, R. A., (1966), "Relation of cue to consequences in avoidance learning", *Psychonomic Science*, **4**, 123–124.

Gibson, J. J., (1979), *The ecological approach to visual perception*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Glass, A. L. & Holyoak, K. J., (1985), *Cognition*, New York: Random House.

Gulz, A., (1991), "The planning of action as a cognitive and biological phenomenon", *Lund University Cognitive Studies*, **2**.

Hebb, D. O., (1949), *The organisation of behavior*, New York: John Wiley and Sons.

Holland, J. H., Holyoak, K. J., Nisbett, R. E. & Thagard, P. R., (1986), *Induction: Processes of inference, learning and discovery*, Cambridge, MA: MIT Press.

Horswill, I. D. & Brooks, R. A., (1988), "Situated vision in a dynamic world: chasing objects". *Proceedings of AAAI–88*, St Paul, MN, 796–800.

Horswill, I., (1992), "Characterizing adaption by constraints", in F. J. Varela and P. Bourgine (eds.) *Towards an practice of autonomous systems*, 58–63, Cambridge, MA: MIT Press.

Hull, C. L., (1934), "The concept of the habit-family hierarchy and maze learning", *Psychological Review*, **41**, 33–52, 134–152.

Hull, C. L., (1943), *Principles of behavior*, New York: Appleton-Century-Crofts.

Hull, C. L., (1952), *A behavior system*, New Haven: Yale University Press.

Ito, M., (1982), "Mechanisms of motor learning", in S. Amari and M. A. Arbib (eds.) *Competition and cooperation in neural nets*, 418–429, Berlin: Springer-Verlag.

Ito, M., (1989), "Long Term Depression", *Annual Review of Neuroscience*, **12**, 85–102.

Johnson, M. H. & Morton, J., (1991), *Biology and cognitive development: the case of face recognition,* Oxford: Blackwell.

Kraft, L. G., III & Campagna, D. P., (1990), "A summary comparison of CMAC neural network and traditional adaptive control systems", in W. T. Miller, III, R. S. Sutton and P. J. Werbos (eds.) *Neural networks for control*, 143–169, Cambridge, MA: MIT Press.

Lippman, R. P., (1987), "An introduction to computing with neural nets", *IEEE ASSP magazine*, 4–22.

Lorenz, K., (1977), *Behind the mirror*, London: Methuen & co Ltd.

Macfarlane, D. A., (1930), "The role of kinesthesis in maze learning", *University of California Publications in Psychology*, **4**, 277–305.

Mackintosh, N. J., (1974), *The psychology of animal learning*, New York: Academic Press.

Mackintosh, N. J., (1983), *Conditioning and associative learning*, Oxford: Oxford University Press.

Maes, P., (1990a), "Designing autonomous agents", in P. Maes (ed.) *Designing autonomous agents*, 1–2, Cambridge, MA: MIT Press.

Maes, P., (1990b), *Designing autonomous agents*, Cambridge, MA: MIT Press.

Marler, P., (1970), A comparative approach to vocal learning: song-development in white-crowned sparrows, *Journal of Comparative and Physiological Psychology Monograph*, 71, pt. 2, 1–25.

Masterson, F. A., (1970), "Is termination of a warning signal an effective reward for the rat?", *Journal of Comparative and Physiological Psychology*, **72**, 471–475.

Melzoff, A. N. & Moore, M. K., (1977), "Imitation of facial and manual gestures by human neonates", *Science*, **198**, 75–78.

Moore, J. W., & Blazis, D. E. J., (1989), "Cerebellar implementation of a computational model of classical conditioning", in Strata, P. (ed.), *The olivocerebellar system in motor control*, Berlin: Springer-Verlag.

Morris, R. G. M., (1981), "Spatial location does not require the presence of local cues", *Learning and Motivation*, **12**, 239–260.

Newell, A., (1990), *Unified theories of cognition*, Cambridge, MA: Harvard University Press.

Olton, D. S. & Samuelson, R. J., (1976), "Remembrance of places past: spatial memory in rats", *Journal of Experimental Psychology: Animal Behaviour processes*, **2**, 97–116.

Pavlov, I. P., (1927), *Conditioned reflexes*, Oxford: Oxford University Press.

Payton, D. W., (1990), "Internalized plans: a representation for action resources", in P. Maes (ed.) *Designing autonomous agents*, 89–103, Cambridge, MA: MIT Press.

Peng, J. & Williams, R. J., (1993), "Efficient learning and planning within the Dyna framework", *Adaptive Behavior*, **1**, 437–454.

Premack, D., (1965), "Reinforcement theory", in D. Levine (ed.) *Nebraska symposium on motivation*, Lincoln: University of Nebraska Press.

Premack, D., (1971), "Catching up on common sense, or two sides of a generalization: reinforcement and punishment", in R. Glaser (ed.) *On the nature of reinforcement*, New York: Academic Press.

Rescorla, R. A. & Wagner, A. R., (1972), "A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement", in A. H. Black and W. F. Prokasy (eds.) *Classical conditioning II: current research and theory*, 64–99, New York: Appleton-Century-Crofts.

Rosch, E., (1973), "On the internal structure of perceptual and semantic categories", *Journal of Experimantal Psychology: General*, **104**, 192–233.

Rosen, R., (1985), *Anticipatory systems – philosophical, mathematical and methodological foundations*, Elmsford, N.Y: Pergamon Press.

Runeson, S., (1989), "A note on the utility of ecological incomplete invariants", *Newsletter of the international society for ecological psychology*, **4**, 6–9.

Samuel, A. L., (1959), "Some studies in machine learning using the game of checkers", *IBM Journal of Research Development*, **3**, 210–229.

Schmajuk, N. A. & Blair, H. T., (1993), "Place learning and the dynamics of spatial navigation: a neural network approach", *Adaptive Behavior*, **1**, 353–385.

Schmajuk, N. A. & Thieme, A. D., (1992), "Purposive behavior and cognitive mapping: a neural network model", *Biological Cybernetics*, **67**, 165–174.

Seligman, M. E. P., (1970), "On the generality of the laws of learning", *Psychological Review*, **77**, 406–418.

Sjölander, S., (1993), "Some cognitive break-through in the evolution of cognition and consciousness, and their impact on the biology of language", *Evolution and Cognition, in press*.

Skarda, C. A. & Freeman, W. J., (1987), "How brains make chaos in order to make sense of the world", *Behavioral and brain sciences*, **10**, 161–195.

Small, W. S., (1901), "Experimental study of the mental processes of the rat II", *American Journal of Psychology*, **12**, 206–239.

Sutton, R. S., (1992), "Reinforcement learning architectures for animats", in J.-A. Meyer and S. W. Wilson (eds.) *From animals to animats*, 288–296, Cambridge, MA: MIT Press.

Sutton, R. S. & Barto, A. G., (1990), "Time-derivative models of Pavlovian reinforcement", in M. Gabriel and J. Moore (eds.) *learning and computational neuroscience: foundations of adaptive networks*, 497–538, Cambridge, MA: MIT Press.

Timberlake, W., (1984), "The functional organization of appetitive behavior: behavior systems and learning", in M. D. Zeiler and P. Harzem (eds.) *Advances in the analysis of behavior*, New York: Wiley.

Tolman, E. C., (1932), *Purposive behaviour in animals and men*, New York: Appleton-Century-Crofts.

Tolman, E. C. & Honzik, C. H., (1930), "'Insight' in rats", *University of California Publications in Psychology*, **4**, 215–232.

Watkins, C. J. C. H., (1992), "Q-learning", *Machine Learning*, **8**, 279–292.

Widrow, B. & Hoff, M. E., (1960/1988), "Adaptive switching circuits", in J. A. Anderson and E. Rosenfeld (eds.) *Neurocomputing: foundations of research*, 123–134, Cambridge, MA: MIT Press.

Zipser, D., (1985), "A computational model of hippocampal place fields", *Behavioral Neuroscience*, **99**, 1006–1018.