Natural Intelligence in Artificial Creatures

Christian Balkenius

Lund University Cognitive Studies 37
1995

First edition

Printed in Sweden by Lunds Offset
ISBN 91-628-1599-7
ISSN 1101-8453
ISRN LUHFDA/HFKO--1004--SE
Lund University Cognitive Studies 37

To my friends and enemies

Table of Contents

1. Introduction 9
2. Biological Learning 15
3. Design Principles 47
4. Reactive Behavior 79
5. Modeling Learning 127
6. Motivation and Emotion 161
7. Perception 185
8. Spatial Orientation 207
9. Cognition 227
10. Conclusion 245

Appendices 261

Chapter 1

Introduction

It is the goal of several sciences to construct models of behavior and cognition. Two fundamental questions for all such endeavors are: What mechanisms are required to support cognitive processes in an animal or a robot. How do such mechanisms interact with each other? This book is an attempt to study these questions within the field of behavior-based systems and artificial neural networks.

The overall task will be to construct complete, artificial nervous systems for simulated artificial creatures. This enterprise will take as its starting point studies made of biological systems within ethology and animal learning theory. We will also consider many ideas from neurobiology and psychology, as well as from behavior-based robotics and control theory. All these areas have valuable insights to contribute to the understanding of cognition.

Ethology has stressed the importance of innate fixed-action patterns, or instincts, in the explanation of behavior. Another significant contribution is the demand that behavior should be studied in the natural habitat of an animal. This leads to a view of behavior and cognition which is very different from the one suggested by animal learning theory. This latter theory attempts to understand the basis of learning by observing the behavior of animals in laboratory experiments. It suggests a view of cognition that is complementary to that offered by ethology, since it stresses the role of learning rather than innate mechanisms. However, as more empirical data become available, the models in both ethology and animal learning theory gradually converge on what may become a substantially more unified theory of animal learning and behavior.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

Other important insights about the mechanisms of cognition are offered by neurobiology and psychology. The goal of neurobiology is to uncover the neurophysiological mechanisms of the nervous system from the neural level and up. Much research within this tradition investigates the properties of individual neurons, however, but here we will mainly consider models at a system level. Such models are closer to psychology where a number of relevant models can be found. In this book, we will consider results from both the behavioral and cognitive traditions in psychology. Although we do not adhere to the views of the behavioral position, much of the terminology we will use originates from that tradition. However, we will mainly look at phenomena typically studied in the more cognitive approach, such as expectancy, categorization, planning and problem solving.

From the engineering sciences we will borrow ideas from both behavior-based robotics and control theory. In many respects, behavior-based robotics is the counterpart of ethology within robotics. Many of the models within this area are very similar to those proposed in ethology. The important difference is, of course, that behavior-based robotics attempts to build working robots, and is not an attempt to study biological systems. A number of concepts from control theory will also be used in this book. The most important one is the view of an animal as engaged in closed-loop interaction with the environment.

Somewhere in the middle of these fields is cognitive science with the ambition to cut across the boundaries of these more traditional approaches (Norman 1990). The present book is such an attempt to combine ideas from all these different areas.

This book has three goals. The first is to identify the systems required in a complete, artificial creature. We will argue that such a creature requires a large set of interacting systems. Some of these are fixed, while others must include different types of learning mechanisms. Our main task will be to identify these systems, rather than give any final solutions to their operation. We will, however, take care to construct fully working miniature models of all the proposed systems.

The second goal is to investigate how the different systems should interact with each other to make the overall behavior of the creature consistent. Many different models have been proposed in the various areas we will consider, and our attempt will be to make an inventory of these different mechanisms. Again, we will propose a number of fully worked out mechanisms.

Finally, we want to map out the way for more cognitive abilities, such as planning and problem solving. We believe that an overall emphasis on the concept of expectations will promote a transition to such abilities.

In taking a design perspective to animal behavior and learning, we will consider how to construct systems that produce sensible coherent behavior rather than try to explain behavioral data. If successful, this approach should give us insights about why real animals are constructed as they are. This requires that the components of the model are developed to a level where they can successfully operate together.

The model proposed here will be based on a large set of findings within animal learning theory, but our goal is not to settle any disputes about animal learning or behavior. Instead, the aim of this book is to construct a set of mechanisms that reflect those found in biological systems. The goal is, thus, to find a consistent model of a complete creature. Since the model we will propose is computational, consistency will always be the prime condition, and agreement with empirical data only a secondary requirement. Of course, this does not mean that we will ignore empirical data, but it will not be ultimately constraining. The creatures developed will, in fact, be mostly based on empirical findings, although it is necessary to simplify many details in order to get the overall system to function.

Even though it would obviously be interesting to try to emulate neurophysiological and behavioral data more closely, the current knowledge of the brain makes such an endeavor very difficult, even for a very restricted sub-system. To construct an entirely realistic model of a complete nervous system based on our current knowledge is clearly impossible. The proposed model can, thus, be compared to real nervous systems on a functional level only. We believe, however, that the functional sub-systems we propose must have parallels in real nervous systems. A complete model of a creature can, therefore, be of great use in two areas.

The first is in the study of biological systems where it can be used both to suggest mechanisms to look for, and to give an understanding of the number of systems interacting with each other. We hope the model proposed in this book will give the overall picture that is often missing when specific abilities or systems are discussed. It should be kept in mind, however, that this book deals primarily with artificial creatures, and as such, it cannot give us any direct model of any particular real animal. Such questions are better handled by empirical studies.

The second area where the model can be used is within autonomous systems. Since the model is detailed enough to be implemented in a computer, it can also potentially be adapted for robotic control. This would very likely require many changes within low level aspects of the model, but the overall structure would be the same. The presented model can, thus, be seen as a framework for an autonomous agent.

Since it is the overall picture that is our interest, we will try to use as few mathematical concepts as possible, in order to make the text more comprehensible. Formal specifications of all systems are given in the appendices, although there is little formal treatment of the model. Such an analysis would, of course, be interesting, but is not the primary goal of this book. The reported simulations will, thus, have to serve both as examples, and as proof of the performance of the system.

Chapter 2 presents an overview of the different problems that have to be solved in order to construct a model of a complete creature. The emphasis will be on various results from animal learning theory. The goal of this chapter is to show that a general learning system is not realistic from a biological perspective. We will argue that biological systems use many interacting systems for different abilities, and the conclusion will be that it is necessary to take this into account if we want to produce an artificial system with similar abilities. This chapter is, thus, intended both as a presentation of the biological background and as an attempt to set the goal for the model we will be developing in the remaining chapters.

Chapter 3 gives a background to the design principles that will be used in the construction of the model. We will briefly review a number of ideas from behavior–based robotics and discuss how they can be used to constrain the design of artificial creatures. It is argued that the basic building block for artificial creatures should be the *behavior module*, which represents a particular mapping from sensors to effectors, that is, a particular control strategy. It is suggested that behavior modules can be combined into hierarchies called *engagement modules*, each of which controls one particular task of the creature. We also introduce the type of artificial neural network that is used for the artificial nervous systems of our creatures. The chapter concludes with a concrete example of an artificial creature which illustrates how neural networks can be used to control a simple body in a simulated environment.

Chapter 4 initiates the development of the model. We present a taxonomy of different reactive behaviors and a number of elementary components that can be used to construct them. We first discuss the directedness of behavior and identify four general categories of behavior. *Appetitive* behavior is directed toward an attractive object or situation. *Aversive* behavior is directed away from negative situations. *Exploratory* behavior is directed toward stimuli that are novel in the environment. Finally, we describe a class of *neutral* behaviors relating to objects that are neither appetitive nor aversive. This classification is a step away from a single hedonic dimension and it gives a richer framework for understanding reactive behavior. It becomes possible to distinguish between active avoidance used for escape, passive avoidance used to inhibit inappropriate behavior, and neutral avoidance used to negotiate obstacles. The new classification also captures the difference between exploratory and appetitive behavior in a natural way. We finally present a number of ways in which behavior modules can be coordinated both sequentially and in parallel. The chapter concludes with an example of an elementary reactive repertoire for our model creature.

Chapter 5 discusses how adaptation can be included within and between engagement modules to coordinate which behavior modules should be activated or inhibited. Starting from the two classical types of learning: instrumental and classical conditioning, we present a new real-time model of conditioning that can be used for both types of learning. The model combines many properties of earlier two-process models of conditioning (Mowrer 1960, Gray 1975, Klopf 1988), but has the additional ability to distinguish between appetitive, aversive, neutral and unknown situations. It can, thus, select between the different types of behaviors described in chapter 4. The model also shares many properties with other reinforcement learning techniques, such as Q-learning (Watkins 1990) and temporal–difference learning (Sutton and Barto 1990). We will describe how the proposed learning system can model a number of experimental situations, including delay and trace conditioning, backward conditioning, extinction, blocking, overshadowing, and higher-order conditioning. We also give a number of examples of how it can be used within an engagement system for appetitive and aversive learning, for sequential behavior chaining, and for the learning of expectations. The general observation will be that learning should be triggered by a mismatch between the expected and actual sensory state of the system.

Apart from learning, behavior is also influenced by motivation. This concept is discussed in chapter 6, where it is identified with a central system for behavior selection. This system is based on the classical notions of drives (as internal needs) and incentives (as external possibilities) (Hull 1952). We identify the classes of *external incentives* which are based on directly perceivable goals and *internal incentives* derived from stimuli that only predict the goal. These incentives can be either primary, that is, innate, or secondary, that is, acquired using the learning mechanisms described in chapter 5. The motivational system combines information about these factors to form a decision about what the creature should do.

We develop a new model of action selection based on the view of the motivational state as a transient representation of the currently most favored behavior rather than as a fixed goal representation. In this model, motivational competition allows the creature to rapidly switch between different engagements while the positive feedback-loop set up by the incentive mechanism avoids behavioral oscillation. Within this framework, it is possible to interpret emotions as states produced by reinforcing stimuli (Rolls 1990). We will argue that motivational states have the function of telling the creature what it should do, while emotions tell the creature what it should have done. The conclusion will be that the concepts of motivation and emotion must play a central role in a cognitive theory.

In chapter 7, we investigate categorical learning and its relation to perception. We will argue that the role of categories is to reduce the complexity of the perceived world by generating orthogonal representations for similar sensory patterns when needed. This allows expectations to be added together in a straight-forward way. As in chapter 5, learning will be driven by different mismatch conditions. We will identify three situations in which it is necessary to construct new categories. In the first situation, none of the existing categories match the external stimulus situation sufficiently well. In the second case, the creature does not receive the expected reward and, thus, needs a better representation of the situation. Finally, there is the more general case when expectations of the environment are not sufficiently fulfilled. It is shown that when this categorization mechanism is added, the creature can learn the higher-order expectations which are required in negative patterning experiments. We also present a simple model for place-approach which can learn a generalization surface around a goal. This surface can in turn be used to guide the locomotion of the creature when the goal is not directly perceivable. This chapter finally describes how exploration can be driven by unfulfilled expectations. Novel and omitted stimuli in the environment trigger exploratory behavior which helps the creature learn about the new state of the world. By using behavior modules for place-approach, the creature can investigate the location where a stimulus used to be before it was removed.

In chapter 8, we will investigate learning and relearning of behavioral sequences using different types of elementary behaviors and learning mechanisms. We show how the systems proposed earlier in the book can be used to solve problems in the spatial domain. Perceptual categorization is combined with behavior chaining to enable the creature to learn a simple maze. A neural network architecture is proposed which uses recurrent expectations to make local choices about what behavior to perform. This system is able to solve shortcut and detour problems and can be said to organize a cognitive map. We finally propose that procedural and expectancy learning is related to the distinction between implicit and explicit memory.

Chapter 9 discusses how the mechanism presented earlier could possibly be extended to handle more advanced cognitive abilities such as multi-modal categorization, association and generalization. We will sketch how these abilities can be used as a basis for an internal environment, which in turn makes planning and problem solving possible. With the proposed extensions of the model, planning and problem solving become truly emergent properties since there is no distinct planning module within the system. The chapter concludes with a brief discussion of the relation between motivation, emotion and planning.

Finally, chapter 10 presents an overview of the proposed model and shows how the various components can interact with each other in different ways. We also discuss the model from an evolutionary perspective and compare the various systems with functions suggested as residing in different areas of the brain. The concluding section discusses some theoretical and practical limitations of the model and presents directions for further research.

Chapter 2

Biological Learning

2.1 Introduction

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 behavior in terms of a large set of interacting learning mechanisms and innate behaviors. 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 this section we will argue that in the light of the evidence from animal learning theory, such a transition is to be highly desired.

For many years, researchers within both animal learning theory and artificial intelligence have been searching for the general laws of learning. We want to propose that such laws cannot be found for the simple reason that they do not exist. Below, we will give a number of different 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 to depend either on an axiomatic system as in the logical tradition (Charniak and McDermott 1985) or to base all intelligence on a simple principle such as chunking (Newell 1990). In both

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

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 we will argue below, will remain so unless artificial intelligence makes the transition into a more diversified view of intelligence.

Before starting, we would like to make a few remarks on our use of the notions of learning and intelligence. Both terms are, of course, exceedingly vague and we will make no attempt to change that situation. We have, nevertheless, some intuitive appreciation of 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 behavior* and that is indeed the reason for its existence in animals and man (Lorenz 1977). However, it may be too restrictive to exclude behavioral changes which are not adaptive. There are in practice, many behavioral 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 behavior that is more likely than not to be adaptive.

Next, we turn to the concept of intelligence. Behavior is usually considered intelligent when it can be seen as adaptive. An animal is considered intelligent when we can see how its behavior 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, we begin to question its intelligence. Eventually, we find out 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 behavior intelligent only when *we* see how that behavior is adaptive for the animal. This is precisely the idea that “intelligence is in the eyes of the beholder” (Brooks 1991a). We should not, however, be tempted to believe that intelligence is *only* in our eyes. If we change the environment of the animal in such a way that its initial behavior is no longer adaptive, we can make an interesting observation. If the animal persists in its original behavior, we no longer consider it intelligent. If it, on the other hand, changes its behavior to adapt it to the new circumstances, we will still think of it as intelligent in some sense. In our opinion, this perspective makes intelligence equivalent to the capacity of learning.

---

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

2.2 The Legacy of Behaviorism

During the reign of behaviorism it was habitually taken for granted that all behavior could be explained in terms of stimulus–response (S–R) associations. Based on this assumption, innumerable experiments were conducted with one single 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 thought, 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 more and more simple until the development culminated in the ingenious Skinner box. This device was entirely devoid of any behavioral 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.

Based on the fundamental belief that all behavior in all species could be explained in terms S-R associations, it was entirely immaterial for the behaviorists whether they 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 whether 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 of 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.

---

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

2.3 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 a 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.

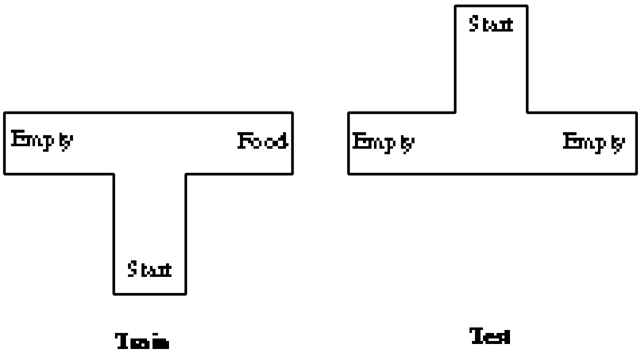
Whatever they had learned, it could not have been the response of performing some specific swimming motion associated with the stimuli at each place in the maze. According to Tolman, the rats had not learned a series of responses but instead the spatial layout of the maze. This 'cognitive map' could 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 2.3.1. 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 to 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. The important observation made is that in *some* cases, a place strategy is clearly within the ability of rats.

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 difference 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 toward 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, the presence of place learning is beyond doubt in the case of Morris' water tank.



Train

Test

Figure 2.3.1 Mazes for a place–learning experiment.

Numerous experiments have been made where 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 type of learning. This is certainly some sort of response learning, but whether a stimulus is involved or not, 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 described by Mackintosh 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 is much more advanced than stimulus–response learning. Perhaps the reason why it has been so hard to find *the* learning mechanism in animals is simply that it does not exist. This would leave us with two possibilities: either there is no learning at all or there are a number of interacting learning mechanisms.

2.4 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 behaviors do not solely depend on this ability. This is what makes cats different from dogs and mice different from men. At this point, we must enter the area of species–specific behavior.

Such behaviors 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 concept was extended to explain all animal behavior 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 behavior. 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 behavior. Perhaps what we believe to be intelligence is only an amalgamation of such fixed behaviors. Can it be that learning only plays the role of adjusting these fixed behaviors 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 which considers the acquisition of grammar as an instance of parameter setting of the above type. Though we 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) found in 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 behavior 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’ behavior of newborn children (Stein and Meredith 1993). Almost immediately after birth, a child will imitate a number of facial gestures such as sticking the tongue out or opening the mouth. 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 shows some degree of adaptation.

A related mechanism is the smiling ‘reflex’ that also can be shown in neonates (Melzoff and Moore 1977). 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 behavior 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 behavior, for example, of 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 to 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).

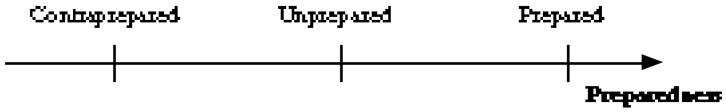


Figure 2.4.1 The Preparedness Dimension

Seligman challenges what he calls the *assumption of equivalence of associability*. This is precisely the assumption that was the driving force behind the behaviorist 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 2.4.1). 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 behavioristic tradition, the animal is typically unprepared for the task. Ethologists, on the other hand, 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 1,000 trials.

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 sick by being injected with a toxic substance. Two other groups of rats were allowed to drink water tasting saccharine. One of these groups was also given electric shock while the other was made sick.

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 to the water. On the other hand, the group that had been made sick did not show any aversion to 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. The group that had been made ill showed an aversion to 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 2.4.2).

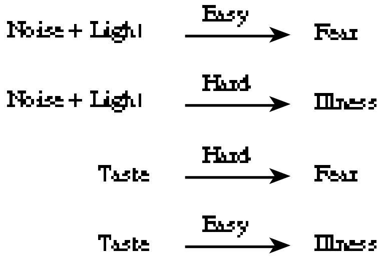


Figure 2.4.2 The potential associations in the Garcia and Koelling experiment.

It has been pointed out that the equivalence of associability is not required by the traditional behaviorist approach (Timberlake 1983). 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 the traditional approach 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 this 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 behaviors such as running are learned almost immediately. In an influential paper on the subject, Bolles (1970) suggested that just like animals have specific behaviors for other engagements such as eating, obtaining food and mating, they must also have innate defence behaviors.

> Such behaviors 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 behaviors 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 behavior and consequently in agreement with food. But this behavior 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 of 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 behaviorists were ever able to teach their animals any of their arbitrary behaviors.

> The truth of the matter is that our finest learning researchers have been keen observers of the organization underlying an animal’s behavior; 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 1983, p. 183)

We may conclude that there exist a large number of innate behaviors which interact with learning in a highly complex way. These innate behaviors may make learning either easier or harder. There also exist innate preferences for forming certain associations and not others. Again we see that there is nothing general about learning. The supposedly general law the behaviorists tried to discover was the result of the arbitrariness of their experiments. In 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.

2.5 Interacting Learning Systems

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

A conclusion that can be drawn from this experiment is that there exist two distinct learning mechanisms which are used in the discrimination task. The first mechanism is concerned with the categorization of odors while the second mechanism is used to associate odor categories with the appropriate responses. Learning by the second system is typically performed on a single trial once the odors are known, while the first system is somewhat slower. This would explain why prior exposure to the odors 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 behaviorists, we find researchers holding 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 conditioning (Hull 1952).

While Hull’s two systems are no longer considered an accurate model of learning, they do show that not all behaviorists believed in one general learning system. It should be noted that Hull was one of the few early psychologists that were more interested in fitting the theory to data than selecting data supporting 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).

2.6 The Role of Reinforcement

We have seen above that learning of odors can occur entirely without reinforcement although this learning may not be expressed in behavior 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 learning was not expressed until reinforcement was introduced. This phenomenon is known as *latent learning*.

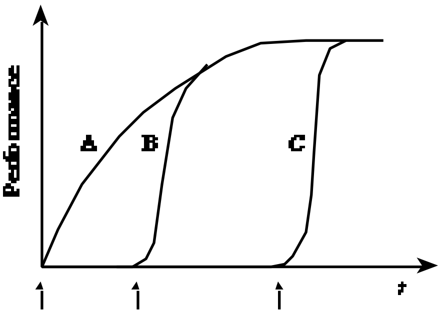


Figure 2.6.1 The idealized learning curves of three groups (A, B, C) in a latent learning experiment. The small arrows show when reward is introduced for each of the groups. When this happens, the performance of each group approaches that of the group which has received reward all along.

Figure 2.6.1 exemplifies the learning curves in an 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, 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.

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 (Hull 1952). 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 binball 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 reward a less probable activity. The exact nature of reinforcement is however still debated and will probably continue to be so for a long time.

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 where the principle does not hold (see Dunham 1977). It appears that reinforcement does play a role in some but not all learning.

A different view of these matters is given by Gallistel (1990), who argues that there need not be any direct relation between the learning situation and the behavioral context in which the animal makes use of the acquired knowledge or habit. For example, certain migratory birds learn the constellations of the stars at a time when they cannot yet fly. Since the stars do not play any role in the nestbound stage of their life, it cannot be the utility of the acquired knowledge that reiniforces learning. This process thus appears to be similar to imprinting. It relies on an innate mechanism which triggers learning under some specific condition. There can obviously be no general principle for this type of learning.

2.7 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 of 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 we will try to summarize in this section. Here we 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 associated with the correct response to perform at that place. A problem with this approach is that the behavior 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 behaviors are *divergent*. As a sequence of S–R associations is performed, the error will accumulate until it drives the animal off course (See figure 2.7.1). A larger set of S–R associations makes the behavior more stable, but it 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.

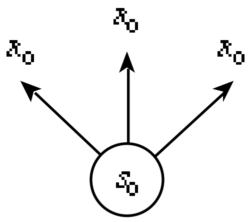


Figure 2.7.1 The result of performing a response depends on the initial position of the animal.

Stimulus–Approach Associations

Another explanation may be that the animal has learned to approach a number of stimuli in the maze. To get to the goal it first has to approach stimulus, S_0, then stimulus, S_1, and so on until it is able to reach the goal box. This behavior can be called stimulus–approach behavior (Schmajuk and Thieme 1992) or beaconing Gallistel (1990). Contrary to stimulus–response behavior, stimulus–approach behavior is stable. This depends on the fact that an approach behavior consists of a whole set of responses 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 2.7.2). As a consequence, stimulus–approach behavior is *convergent*. This makes this type of learning much more likely as a basis for adaptive behavior.

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 which 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.

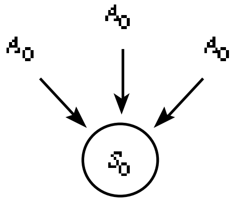


Figure 2.7.2 The stimulus S_0 can be approached from many directions.

Place–Approach Associations

Like stimulus–approach, place–approach 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 previous since it requires the ability to use a configuration of stimuli to identify a place — a far from trivial task. A number of models of this process have been suggested, however. (See for example Schmajuk and Blair 1993 and Zipser 1985.) Figure 2.7.3 shows a place-approach situation. This is the second of the possible mechanisms used in the T-maze discussed above. A behavior of this type may alternatively be called piloting (Gallistel 1990).

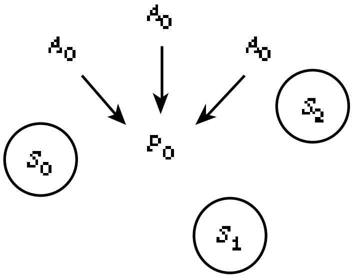


Figure 2.7.3 A place, P_0, defined by a set of stimuli (or landmarks), S_0, S_1, S_2, can be approached from a number of directions.

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 2.7.4). Obviously, this type of behavior is even more unstable than a simple S–R reaction. The use of response chains depends on a very accurate motor system and one would think that learning of this type would not be used, if an animal could choose another strategy.

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.

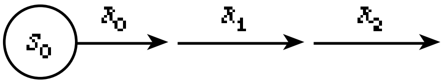


Figure 2.7.4 A Response Chain.

Surprisingly, it appears that animals do use response chains to a much larger extent than could be expected. This is especially the case when they are overtrained on a task. With increased training, the animal is less likely to use stimuli to guide its behavior (Gallistel 1990). Response chains are also used in fixed–action patterns as could be seen above.

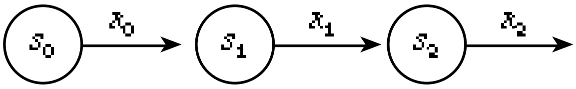


Figure 2.7.5 A S–R Chain.

Chains

Another possibility is that the animal learns stimulus–response chains. These can be seen as pairs of a stimulus and a response, <S_i, R_i>, where response R_i should be executed whenever stimulus, S_i, is present. This response leads to a new stimulus S_j which will trigger the next S–R associations (figure 2.7.5). This is the type of structure which has traditionally been studied within the behavioristic tradition.

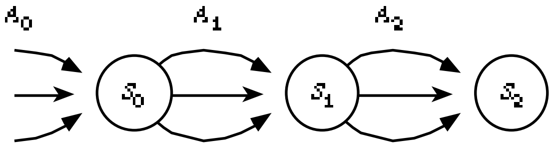


Figure 2.7.6 An Approach Chain.

Stimulus–Approach Chains

Just as responses can be linked together in chains, it is also possible for approach behaviors to be linked. Like a simple stimulus–approach behavior, these chains produce stable behavior, but they can range over much larger distances than a simple stimulus–approach association (figure 2.7.6).

Place–Approach Chains

Naturally, place–approach associations can also be linked in chains (figure 2.7.7). 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.

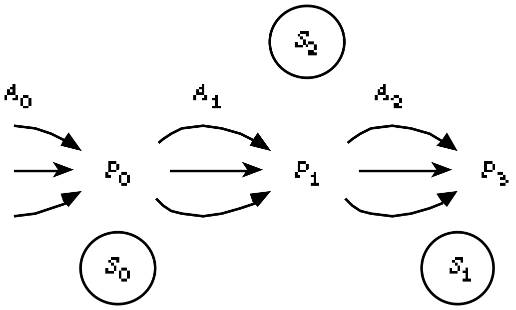


Figure 2.7.7 A Place Chain.

S–R–S' Associations

The types of associations described above can be used to control behavior, but they cannot be used to make inferences. Tolman postulated that animals learn something like S–R–S' associations (Tolman 1932). These tell the animal that if it is in situation S and performs response R, it will end up in situation S'. Associations of this type usually go by the name of *expectancies*. 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 2.7.8). This map can be used with great utility in shortcut and detour problems as well as for general problem solving. It can also be used to detect when a response does not result in the expected situation.

This type 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.

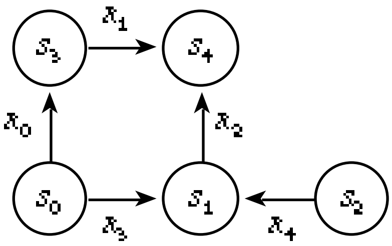


Figure 2.7.8 A map consisting of S–R–S' associations.

A particularly important class of systems can be constructed, if we embed stimulus–approach structures within S–R–S' associations. The behaviors generated by a system of this kind are stable while, at the same time, supporting various forms of inferences. Systems as this 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 learned. As we will see below in chapter 8, 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. We 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 we are advocating is that all these learning types, and perhaps many more, co-exist and interact with each other during learning and behavior.

2.8 Internal Influences on Behavior

So far, we have described behavior as if it were guided primarily by external stimuli. This is of course not the case. Internal determinants of behavior 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 behavior is not directly relevant to the present argument, we will not dwell on this matter here. We have so far assumed that there is only one goal to pursue and that the animal is motivated to do so.

A determinant that is more relevant to the present argument is what we 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 behavior, the rat receives a reward on its first visit to each arm (Olton and Samuelson 1976). Each time the rat is in the center of the maze, it has to choose a new arm to visit (figure 2.8.1). Since the rat cannot perceive the reward from the center of the maze, this behavior seems to require some memory for the previously made choices.

Rats are surprisingly good at this task and they 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.

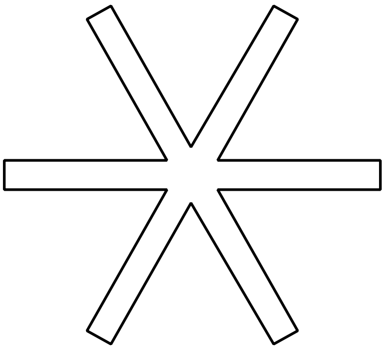


Figure 2.8.1 A radial maze with six identical arms. The animal is rewarded on the first visit to each arm. To be successful, the animal must remember which arms it has already visited.

As a determinant of behavior, the internal context is no different from external stimuli. It is used to direct behavior 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 we have presented above.

2.9 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 which 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 it 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 make it behave in the appropriate way. They do not necessarily need to represent the world in any great detail (Sjölander 1993).

In ecological optics (Gibson 1979), the idea of an *ambient optic array* is used in a way that is very similar to the *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.

---

[3] Umwelt means approximately surrounding environment.

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), 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 behaviorist 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 which were relevant to 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 which 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 behavioral 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. A structure of this kind is much more useful than an objective representation, if the animal wants to move from one place to another.

2.10 When Does the Animal Learn?

In this section we will consider how different proposed learning mechanisms relate to the execution of a *consummatory* or *terminal* behavior. Learning has been described as occurring either before, after, or at the same time as the terminal behavior. We will call these different learning types *early*, *synchronous* and *late* learning (See figure 2.10.1).

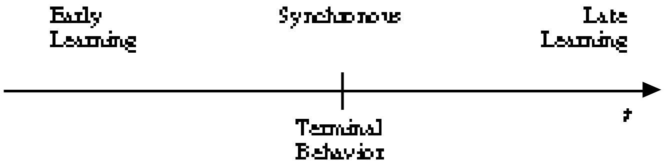


Figure 2.10.1 Different times when learning can occur in relation to a terminal behavior.

Early Learning

Early learning is learning that occurs prior to the consummatory behavior. 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 which 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 which 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 we 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 behaviors, 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 (see Bolles 1978).

It does not appear that synchronous learning can explain the more complex behaviors 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, we 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 behavior 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.

2.11 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 behaviors are necessary to keep the animal alive while it learns about the world. In some animals, almost all behaviors 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 behavior.
- 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.

2.12 Parallels Between Artificial Intelligence and Animal Learning Theory

It is interesting to see that many artificial intelligence models show striking similarities to 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 (Sutton and Barto 1990). The parallel of Tolman's 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 behavior in the work of Brooks and others who emphasize the role of essentially fixed behavioral repertoires which 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? We 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.

2.13 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 learning 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 (Newell 1990). 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 (Albus 1975). 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 which 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: behavior 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 (for example, Ito 1989, Moore and Blazis 1989). Another connection between animal learning theory and control theory is 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).

2.14 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 behavior that precedes it. As a consequence, this behavior is more likely to be elicited again. If, on the other hand, a behavior is followed by some unpleasant event, the likelihood of the behavior 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 Samuel's 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 toward earlier positions of the game. Moves which lead to favorable positions receive a higher reward (that is a better evaluation) than moves which 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 generalized 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 reinforcement learning algorithm that is easiest 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 which 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 it is one of the most recent attempts to propose an explicit connection between animal learning and control theory. Baird and Klopf (1993) describe a modified version of the Q-learning paradigm which also clarifies this connection. They show how Q-learning can be adapted to conform with the precise details of several animal learning experiments.

2.15 Learning as Knowledge Acquisition

According to Tolman, learning is the acquisition of knowledge about the world. This view is the most popular among 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 which seem 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 by Tolman and Honzik (1930), a rat is allowed to explore the maze shown in figure 2.15.1. 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 necessary. 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.

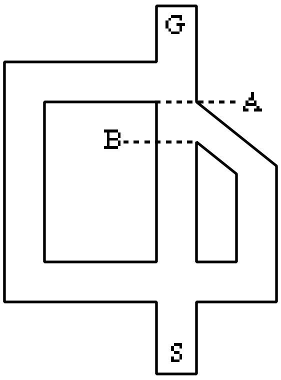


Figure 2.15.1 The maze of the type used in the Tolman and Honzik (1930) experiment.

Tolman's view that learning is essentially the acquisition of knowledge about the environment has no problem explaining this behavior, 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:

$$\text{precondition} \times \text{action} \rightarrow \text{outcome}.$$

A plan is simply a sequence of these representations 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 which leads from the start to the goal. If such a sequence is found, it is subsequently executed in the real world (Newel and Simon 1963).

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 additional experience.

The view that all behavior can be described in this way have received much criticism in recent years and many of the deficiencies of these types of mechanisms have been acknowledged (for example, Maes 1990). 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 take precedence of 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 behavior of the agent. For example, an animal that anticipates its own future needs may gather food even before it becomes hungry (compare Gulz 1991).

In summary, most behavior 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.

2.16 The Reactive Approach

The view that animal behavior 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 1990). The most important aspect of such architectures is their emphasis on complete creatures or systems that let us make observations which cannot be made from studies of isolated modules (Brooks 1986, 1991a, 1991b).

The subsumption architecture introduced by Brooks (1986) is a computational model which 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. The meanings of the messages are given by the operations of both the sender and the receiver (Brooks 1986). 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 behavior based (Connel 1990, Horswill and Brooks 1988). In this context, a behavior 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 behaviors 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 1991a). 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 in situations like the radial maze. It is therefore reassuring to see that Brooks now acknowledges this need (Brooks and Stein 1993).

2.17 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 behaviorist 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 2.17.1. Figure a shows instances of three categories with a discrimination border drawn between them and figure b shows the three categories as bounded regions around the examples.

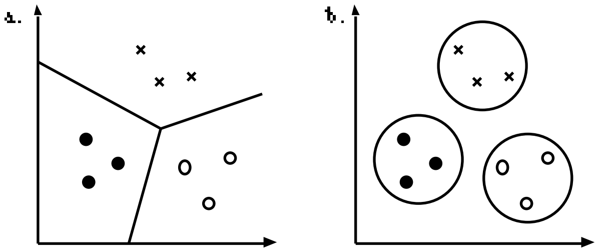


Figure 2.17.1 Two views of categories. (a) Discrimination surfaces in the instance space. (b) Categories as regions.

The main difference between the cognitive and the behavioral approaches does not concern the phenomena studied but rather the way these phenomena are attributed to different mechanisms. The cognitive investigators search for the internal representations of categories while the behaviorists study 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 behavioristic tradition. When the tendency to respond to a stimulus is measured, it can usually be shown that there exists one specific stimulus for which the response is the strongest or the 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 to 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 others can be found both within the area of machine learning (see Davidsson 1994) as well as in neural networks and statistical inference (Lippman 1987).

2.18 Conclusion

All the studies presented above, both within animal learning theory and artificial intelligence, have been concerned with some particular aspect of learning or behavior. To date, very few models have attempted to deal with the full complexity of learning, although there certainly exist biological models which 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 searching for the similarities between the different theories. 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 which have caught most previous attempts.

The first trap is to believe that all learning and behavior can be explained with a small set of principles. The result of this approach has been the construction of grand theories which set out to explain all instances of learning but later are 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 researchers 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.[4]

4. They may, of course, still be useful in technical applications.

In summary, what is needed is an approach where all the different aspects of learning can be combined in an elegant manner. We 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 be 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 (Harnad 1990), 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 vehicle. 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, we suggest that artificial intelligence learns 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. The next chapter is a brief introduction to this area.

Chapter 3

Design Principles

3.1 Introduction

As described in the preceding chapter, there are many ideas from animal learning theory and artificial intelligence that can be used as a starting point for the construction of artificial creatures. In this chapter, we will introduce two areas of research that will play an important role in the rest of the book.

First we will take a look at a number of design principles that have recently been proposed in the field of behavior-based robotics. Systems designed according to these ideas go by the names of situated agents (Agre and Chapman 1987), embedded systems (Kaebling 1993), autonomous agents (Maes 1990), and reactive control (Arkin 1990). The principles introduced here will guide the construction of our artificial creatures. We will also introduce some terminology that will be used in later chapters.

Next, the basic principles behind neural networks will be described with the emphasis on the types of systems that will be used in our artificial creatures. We will see how model neurons can be connected in different ways to perform various functions which will be used later on in the book.

The chapter concludes with a simple example of an artificial creature. In this example, we show how the proposed design principles constrain the construction of an artificial nervous systems. The artificial world that will be used throughout the book is also introduced together with the sensors and effectors of the creature.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

3.2 Behavior-Based Control

A *behavior module* is a subsystem that is responsible for one specific coupling between sensors and actuators (figure 3.2.1). A behavior module can, thus, be thought of as a transfer function which transforms its sensory input to an actuator output. In most of the literature on behavior-based control, a behavior module is referred to simply as a behavior, but since this terminology is rather confusing as well as in conflict with its use in biology and psychology, we will use the longer term instead.

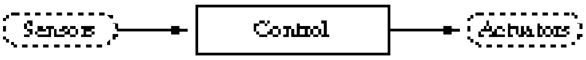


Figure 3.2.1 A behavior module is defined as a transformation from sensory input and to actuator output.

The use of behavior modules as the basic building block contrasts sharply with the view of traditional AI where control is typically based on a set of goals, a model of the world and a search procedure (Brooks 1991a). The search procedure tries to find an action sequence that changes the state of the world to the desired goal state. If such an action sequence is found, it will be executed step by step in the real world.

In a behavior-based agent, the goal needs not be explicitly represented. Instead, behavior modules are selected on an immediate sensory basis in such a way that they are likely to move the agent closer to the goal in the real world. Problems are avoided when they occur. As we will see in later chapters, a behavior-based agent can be augmented with explicit goal representations and planning, but such abilities are not part of its primary repertoire.

The main argument in favor of behavior-based control is that it does not rely on an accurate internal world model. This avoids problems that occur when the internal world model does not agree with the real world. Instead the world is used as its own model in the sense that all actions performed are triggered by the real world as it presents itself at the sensors (Brooks 1990, 1991a, Agre and Chapman 1987). Since the real world is always in agreement with itself, the agent is more likely to react correctly. However, since no internal world model is used, a strictly behavior-based system will not be able to use explicitly represented goals or plans.

Another important aspect of behavior-based systems are thatthey are grounded. This means that they satisfies the *physical grounding hypothesis* that states that to build an intelligent system it is necessary to have its representations grounded in the physical world (Brooks 1991a, Harnad 1990).

The identification between internal behavior modules and external behaviors is somewhat simplified but it expresses the spirit in which our creatures will be built. We will see that behavior in general is an *emergent* function of a system. That is, that there is no strict correspondence between externally observed behavior and internal modules. In the field of behavior-based robotics, emergent functionality is usually though of as a result of the interaction between the agent and the environment. In a more general setting, emergent properties are such that they can not be predicted from the parts that constitute the system (Churchland 1984). The behavior of our creatures will be emergent in both senses of the word.

A single behavior module would not be of much use if it were to operate on its own. To construct a system with some level of sophistication, a number of behavior modules must be combined. There seems to be two distinct ways of doing this. The first is borrowed from the *subsumption* paradigm (Brooks 1986, 1991a, b) and the second has its inspiration in the idea of a central motivational state.

Subsumption

In the subsumption paradigm the control system of an agent consists of a number of behavior modules arranged in a hierarchy. The different layers in the architecture take care of different behaviors. The lower layers control the most basic behaviors of the creature while the higher behavior modules control more advanced functions (figure 3.2.2). This idea is not entirely unlike hierarchical motor patterns as they are used in ethology (Tinbergen 1951/1989, Tyrell 1993).

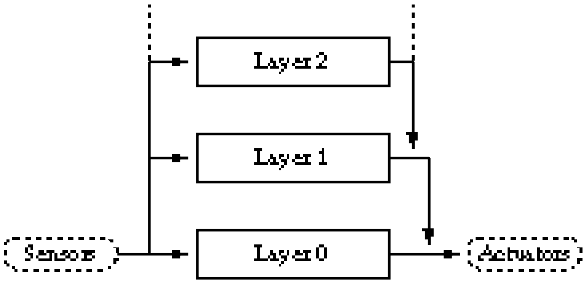


Figure 3.2.2 A hierarchy of behavior modules.

A typical low level behavior in a subsumption style creature controls activities such as object avoidance, wandering and exploration. On a higher level we may find processes like object identification and planning. Each higher layer is able to monitor and control the underlying layers, but the communication between modules is reduced to a minimum. While traditional systems can be said to be vertically decomposed into processing stages, a system of the present type is horizontally decomposed[1] into behavior modules (Schnepf 1991).

---

[1] In practice, the hierarchy is very seldom strict, which means that it is somewhat erroneous to talk a bout a subsumption hierarchy.

An important type of subsumption hierarchy can be called an appetence system. It consists of at least two parts that control *appetence* and *consummatory* behavior. Most parallel engagements can be divided into these two components. For the eating behavior, the appetence behavior consists of searching for, or collecting food, while the consummatory, or *terminal*, behavior corresponds to the actual eating of the food (Lorenz 1977). The distinction between the two is that the first behavior is instrumental in achieving the second. The appetence and the consummatory behaviors can be organized in a subsumption style hierarchy (figure 3.2.3).

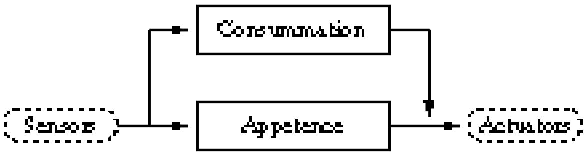


Figure 3.2.3 An appetence system consists of behavior modules for appetence and consummation.

The consummatory behavior, such as eating or sex, is generally very rigid while the appetence behavior can gain a lot from learning. Most of the learning processes that we will investigate in chapter 5 will operate on this part of the behavior.

A glance at the animal learning literature tells us that this basic appetence system is usually much more elaborated in real animals. For example, Timberlake (1983) suggests that the appetitive behaviors of animals are controlled by what he calls *behavior systems*.[2] There exist different behavior systems that are related to feeding, mating, parenting, bodycare and other activities. In relation to learning, it is important to understand that the learning methods involved in the different parts of a behavior system are very different.

For feeding, the behavior system consists of behaviors for individual foraging, social–approach, investigation, predation, food handling, hoarding, ingestion and possibly rejection of the food. All these behaviors can be further divided into smaller pieces. For example, investigation may consist of approach, sniff, nose, lick and bite (Timberlake 1983). Clearly this view of behavior can easily be reconciled with ideas from the subsumption paradigm (see also Collier 1983).

Parallel Engagements

In our artificial creatures, the parallel to behavior systems in real animals will be *engagement modules*. Different engagements are controlled by parallel systems. In an artificial creature, behaviors such as eating and sleeping may be implemented

---

[2] Timberlake's (1983) use of the term 'behavior system' should not be confused with that of Hull (1952). In Timberlake's use of the term, behavior system refers to a subsystem while in Hull's terminology it refers to a description of a complete system.

as parallel engagements. Each engagement is controlled by its own subsumption hierarchy as shown in figure 3.2.4.

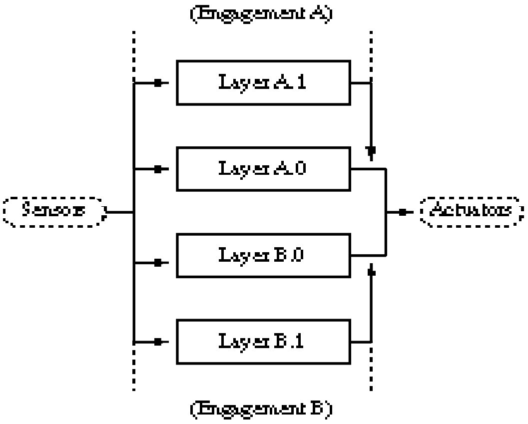


Figure 3.2.4 Each engagement has its own hierarchy of behavior modules.

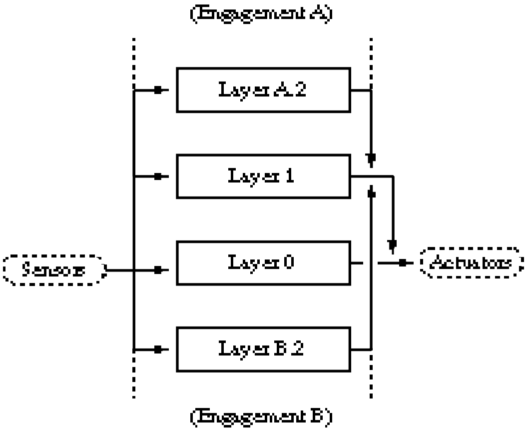


Figure 3.2.5 Two parallel engagement systems that share two behavior modules.

Since some layers of different engagement modules will usually be the same, it is possible for different engagement modules to share behavior modules. Figure 3.2.5 shows an architecture where the two layers are shared by two parallel engagement modules.

It is clear that rather complex systems can be built by combining behavior modules into larger structures in the two ways that we have described so far. When a single hierarchy is used, the subsumption mechanisms are sufficient to choose between the different behavior modules, but when parallel engagements are introduced, some additional mechanism is necessary to restrain different engagement hierarchies from interfering with each other. This is the role of central engagement selection.

Central Engagement Selection

When an agent has different and competing engagements, it must be able to select among these activities in a rational way. To do this, it must evaluate its needs and its possibilities and make a decision about what to do. This decision is represented by a functionally central motivational state that is responsible for activating or inhibiting the appropriate modules for the selected engagement (Gallistel 1980). We will call the part of the agent that handles this decision the *motivational module*.

It has sometimes been argued that a central motivational state is neither necessary nor very useful in an autonomous agent (Maes 1991). It seems, however, that the problem with central motivational states lies, not in making a central decision, but rather in the view that once a decision is made, it is not changed until its corresponding goal is reached.

The type of motivational state we are advocating is continuously reevaluated and does, thus, represent a transient momentary decision and not a long term commitment. This motivational state can change at any moment to reflect changes in the environment or a new evaluation of old knowledge (Balkenius 1993)

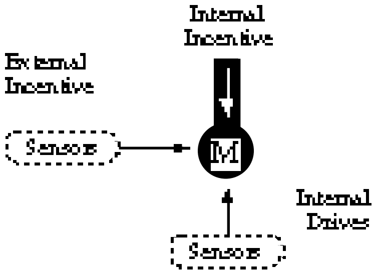


Figure 3.2.6 The determinants of motivation. The motivational state is a function of three factors, internal drives tells the creature about its current needs. External incentive tells it about goal objects which are directly accessible, and internal incentive tells it about more distant possibilities.

Nor is it necessary that the motivational module is physically located at one particular place in the agent. It needs to be central only in the sense that the decision made has global consequences for the whole creature. It is important not to misunderstand the principle of central engagement selection. It simply states that all behaviors cannot be executed at the same time and that the choice of engagement cannot be made locally. One part of the agent cannot decide to search for food while another decides to dance. However, most engagements, and especially parts of behaviors, are best handled in a distributed fashion using a subsumption architecture. The motivational state is selected as a result of three factors (figure 3.2.6).

- **Internal Drives** The creature has a number of primitive needs that vary dynamically over time. When a need is not fulfilled, an internal drive signal is generated that increases the probability of the agent selecting actions that serve to fulfil that need. For instance, one drive could correspond to the need to eat while another drive could make the animal look for predators at regular intervals.
- **External Incentive** At all times, the agent receives sensory input that tells it about the possibility of fulfilling a need. For instance, viewing or smelling food would constitute an incentive to eat. External incentive is, thus, a representation of the possibilities of the immediate environment given by perceivable goal objects such as food.
- **Internal Incentive** Internal incentive plays the same role as primary except that it does not directly depend on the currently perceived situation. Instead it is generated by some internal process as a result of prior learning. For example, an expected food situation would make the agent more likely to search for food even if the food is not in sight. Another example is an open door that is known to lead to food although the food in itself is not visible.

It is important to recognize that the difference between external and internal incentive is not whether learning is involved or not. Both processes can depend on learning. For instance, learning that an apple tastes good could make the apple an external incentive at a later stage. Instead it is the presence of the goal object that determines whether we have a case of external or internal incentive.

The three factors are weighted together for each of the engagements of the agent and a decision is formed about what the agent should do. It is quite possible that this decision is changed as the agent tries to pursue its goal. If our creature comes close to water on its way to the food, it may very well change its decision and drink on its way. Once the need for water has decreased, the creature will continue toward the food. This is an example of opportunistic behavior that results from the interaction between a motivational module and learning or a reactive control system. Figure 3.2.7 shows an architecture where two engagements, A and B, centrally compete for activation.

Apart from determine what the agent should do, the motivational system plays an important role in directing learning. Since the motivational module is the only subsystem that contains information about what the agent is trying to do, it is in a favorable position to determine the success of the currently generated behavior. When the actual success of a behavior does not match the expectations, the motivational system can generate learning signals to the other parts of the creature. As we will argue in chapter 6, such learning signals make up the basis for *emotions*. There, we will also return to the role of the motivational module in learning and give a longer argument in favor of its inclusion in artificial creatures.

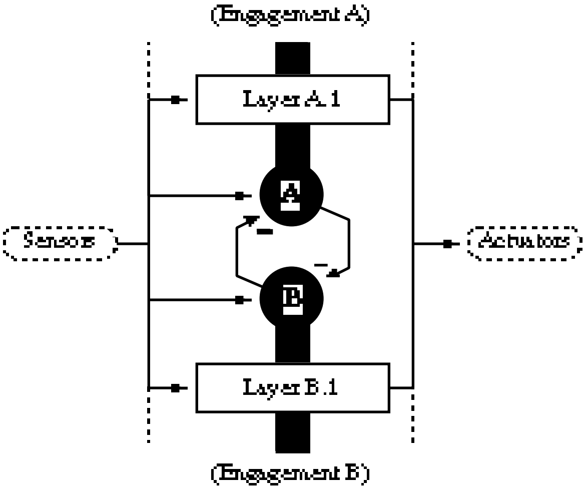


Figure 3.2.7 A behavioral hierarchy is selected centrally.

Functional Levels

The behavior generated by an animal has its causes both within the animal and externally. Let us introduce a useful distinction between three levels of internal causes of behavior (figure 3.2.8, see also Balkenius 1994b and Gärdenfors and Balkenius 1993). A less sophisticated creature may have only the first level while a more advanced animal may have all three.

The first level refers to the innate properties of the system. At this level we find the fixed motor patterns of many animals and the basic reactive behavior of our artificial creatures. At the second level, the innate behavior becomes adaptable by the introduction of learning. Such learning is a direct consequence of experience with the world. The final level is engaged in self-supervised learning. This type of learning is not governed directly by the external world. Instead it is controlled by knowledge of the external world and the consequences of actions performed in it. Adaptation based on self-knowledge is a special type of learning at this level. In chapter 9, we will return to this last level.

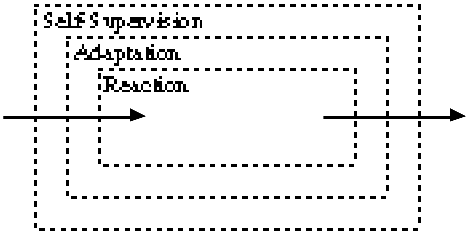


Figure 3.2.8 Three functionals levels of behavior control.

It is not necessary that the three levels correspond to physically distinct modules. For instance, in a module built as a neural network with learning nodes, it is not possible to physically distinguish between the innate and the learning level. At a functional level, however, the two levels are still different. We have not been able to find any example of a system where all three levels exist in a single component, but they probably exist.

When constructing artificial creatures, we will find it necessary to incorporate the three functional levels both within small subsystems such as a single neuron and at a larger scale where the different levels are handled by distinct modules.

Summary

We have presented a number of architectural principles which are useful as design directions when constructing an artificial creature. The most important principle is that autonomous agents should be behavior-based. Once this principle is established we have to consider how to combine behavior modules into larger systems. We have proposed two possible mechanisms. The first is the subsumption architecture introduced by Brooks (1986) and the second is a central motivational system.

In whatever way the behavior modules are combined, the generated behavior can be caused at three different functional levels. The first level corresponds to innate behavior, the second to learned behavior and the third to behavior which has been rehearsed internally or is governed by self-knowledge of some kind.

The ideas presented are mainly based on the type of mechanisms one would find in biological systems. We believe that the type of overall architecture presented here can be used as a starting point for the development of very advanced and capable autonomous agent. In section 1.5, we will consider a very simple creature constructed according to these principles, but first we need to consider some properties of the artificial neural networks which will be used for the nervous systems of our creatures.

3.3 Artificial Neural Networks: A Brief Introduction

Artificial neural networks are, of course, the obvious choice for the nervous system of an artificial creature. However, there are more reasons than simply the name. The most important is that by using artificial neural networks, the representations used will be of the kind used in real brains. These representations use *labeled line coding* with graded activity values (Martin 1991). The idea behind labeled line coding is that each line, in this case, the connection from one node to another, has a fixed meaning. This type of representation is easiest to understand in the case of sensory representations. A specific signal means that a specific sensory stimulus is present. The intensity of the signal reflects directly the intensity of the stimulus. The use of labeled line coding does not imply that representations must be *local*, that is, it is not necessary that one single line codes for one single stimulus. In most cases, a large set of lines will be active for any particular stimulus. This is called a *distributed* representation. Such representations can be comfortably represented as vectors. We do not want to claim that these representations are the same as those in the brain, but they are certainly similar in many respects. While it would be possible to use this type of representation without the neural networks, we think it would be harder to do so consistently.

The next reason for using artificial neural network is that the computations will be "in the style of brain" (Arbib, Conklin and Hill 1987). Again, it is not possible to claim that the computations in the brain are the same, but they are probably more similar to those of neural networks than to any other type of computational model. In this context, we also want to mention that we will make no attempt to make our neurons very realistic. Given the complexity of the model we will develop, and the amount of speculation that inevitably will go into it, it would be nearly impossible to try to be true to the detailed facts about neuronal functioning.

When we will point out the similarities between an artificial neural circuit and some region of the real brain, this will only be a statement about the function they fulfil, not about a similarity at a neuronal or network level. Having said this, we can now safely introduce the neurons and connections which will be used in this book.

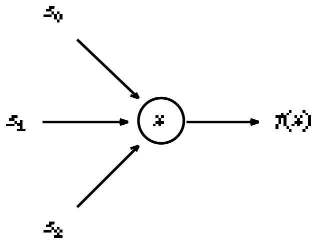


Figure 3.3.1 The neuron model. The node receives three input signals $s_0 \ldots s_2$ that generates an activity x, which in turn controls the output signal $f(x)$.

A Model Neuron

The model neurons we will use in our artificial creatures can be described by two factors. The first is how they combine their *input signals* to form its *activity level*, and the other is how they generate an *output signal* from this activity level (figure 3.3.1).

The input signals will be combined using both temporal and spatial integration. Let $s_0 \ldots s_n$ be the input signals received by a node x. The activity x changes according to differential equation,

$$\frac{dx}{dt} = -\alpha x + \sum_{i=0}^{n} s_i + \alpha R \tag{3.3.1}$$

A node of this type is usually called a *leaky integrator* since it integrates its input over time and leaks through the term $-\alpha x$. If $\alpha=0$, the node does not leak and is simply called an *integrator*. The term αR describes the resting activity of the node. When no input is received, the node will return to this value. Throughout this book, it will be assumed that $R=0$ when it is not explicitly mentioned that it has a different value.

When the input signals changes slowly compared to the time constant of equation (3.3.1), the activity of the node can be approximated by,

$$x = \sum_{i=0}^{n} s_i + R \tag{3.3.2}$$

In this case, the neuron is called a *summing node*. When the integrating property of the node is not important, we will use this type of node since it is easier to analyse. Unless it is explicitly mentioned that integrator nodes are used below, all nodes are assumed to be of the summing type. The output signal from a node is calculated by its output function, $f(x)$. We will use four different types of output functions in this book. The first, and simplest, is the *linear* function,

$$f(x) = x \tag{3.3.3}$$

The second output function is called *semi-linear* (Amari 1982) and includes a threshold level below which no output is generated,

$$f(x) = \begin{cases} x-\theta, & x \geq \theta \\ 0, & x < \theta \end{cases} \tag{3.3.4}$$

Our next output function is the *threshold* function,

$$f(x) = \begin{cases} 1, & x \geq \theta \\ 0, & x < \theta \end{cases} \tag{3.3.5}$$

And finally, we will use a *sigmoid* output function (Cohen 1983),

$$f(x) = \begin{cases} \dfrac{(x-\theta)^n}{\delta + (x-\theta)^n}, & x \geq \theta \\ 0, & x < \theta \end{cases} \tag{3.3.6}$$

Depending on the parameters n and δ, this function can look either as a semi-linear or a threshold function. In either case, it is continuous which will be important in some cases. In the neural network literature, the *logistic* function $f(x) = 1/(1+e^{-x+\theta})$ is often used as sigmoid function (for example, Hopfield 1984), but for the purposes of this book equation (3.3.8) has more suitable properties.

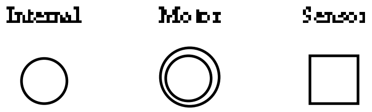


Figure 3.3.2 The Different Node Types.

There are three distinct types of neurons, or nodes, in the nervous systems of our creatures (figure 3.3.2). The first type is the internal node. These nodes are drawn as circles. The second type are the motor, or output, nodes which are drawn as double circles. These nodes control the behavior of the creature. The last type of node is the input, or sensor, node and these are drawn as squares. The output from a sensor node is set by a sensor on the body of a creature (See section 3.4).

The nodes can be connected together by various types of connections. These are shown in figure 3.3.3. When a signal passes through a connection, it is multiplied with the weight, or strength, of the connection. Weights can be either positive or negative. Connections with positive weights are called *excitatory* while connections with negative connections are called *inhibitory*. The excitatory connections are drawn in white while the inhibitory are black.

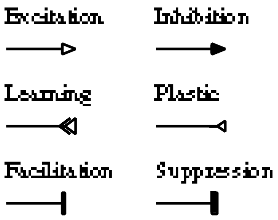


Figure 3.3.3 The Different Connection Types.

If $w_0 \ldots w_n$ are the connection weights on the connection leading to node x, the summed input is given by,

$$\sum_{i=0}^{n} s_i w_i \qquad (3.3.7)$$

Connections can be either *fixed* or *plastic*. Plastic weights change when the node to which they are connected receives a learning signal through a learning connection (figure 3.3.4). The change in the plastic connection, *w*, is governed by a learning rule of the form,

$$\frac{dw}{dt} = h(w, L(x), f(y), g(z)), \tag{3.3.8}$$

where $L(x)$ is the learning signal transmitted from node *x*, $f(y)$ is the output of node *y*, and $g(z)$ is the signal passing through the plastic connection. Various learning rules will be presented as they are needed.

Another type of connection is the facilitating one (figure 3.3.4 c and d). Instead of influencing the activity of another node, a facilitating signal changes the signal in an other connection. Let *F* be a facilitating *signal* and $f(x)$ be the signal in the facilitated connection from node *x* to node y with the weight *w*. The resulting signal is the product of the two signals, that is $Ff(x)$. When this signal reaches node y it will also have been multiplied with the connection strength w. The signal reaching y will, thus, be $Ff(x)w$.

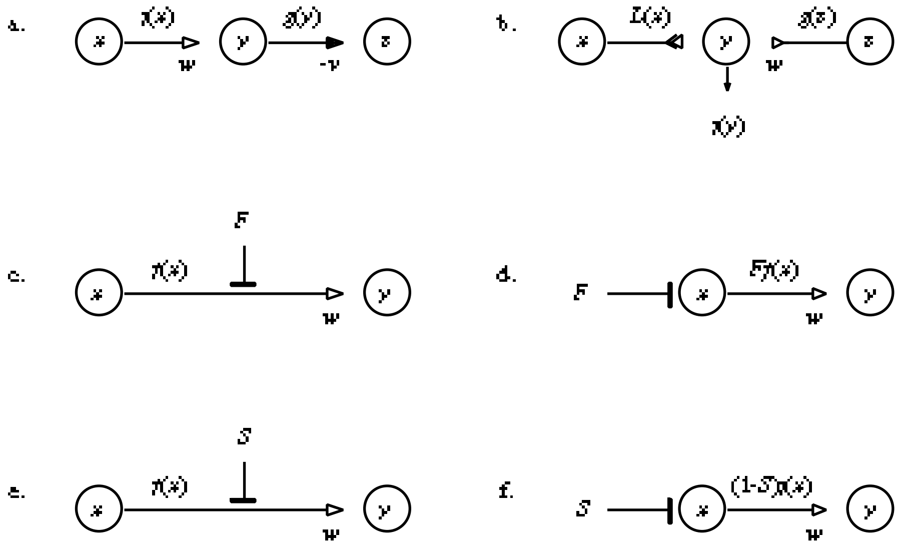


Figure 3.3.4 Examples of connections. (a) The node *x* excites the node *y*. This node in turn inhibits *z*. (b) The node *x* sends a learning signal to *y* which influences the plastic connection, *w*, between *z* and *y*. (c) The *signal* sent from *x* to *y* is facilitated (or gated) by the signal *F*. (d) The *output* from node *x* is facilitated (or gated) by the signal *F*. (e) The *signal* sent from *x* to *y* is suppressed by the signal *S*. (f) The *output* from node *x* is suppressed by the signal *S*.

It is also possible for a facilitating connection to facilitate a whole node and not only a connection (figure 3.3.4 d). In this case, it is the *output* of the node that is multiplied by the facilitating signal. In the example in figure 3.3.4 this gives the same result as facilitation of a single connection, but it is not necessarily the case.

The negative counterpart of facilitation is suppression. This situation is similar to facilitation except that the suppressed signal from x to y is multiplied with $(1–S)$ instead, that is, if $f(x)$ is the signal transmitted in the suppressed connection and w is the connection strength, the resulting signal will be $(1–S)f(x)w$. This means that the larger the suppressing signal is, the smaller the suppressed signal will get. See figure 3.3.4 e and f. The behavior of facilitated or suppressed nodes and connections are similar to that of sigma–pi units where inputs to a node can interact both additively and multiplicatively (Williams 1986).

A final property of connections is that they delay the signal transmitted in them. We will assume that all connections have a fixed time delay of 1 time unit. This means that if node x sends the signal $f(x)$ to node y through a connection with weight w at time t, the activity of y will be increased by $f(x)w$ at time $t+1$. Taking this into account we can summarize the calculation of the activities of the nervous system as follows.

Let $x_0 \ldots x_n$ be the activities of all the nodes in the network and let $f_i(x_i)$ be their output functions. The connection from node x_i to x_j is called w_{ij}. When only summing nodes are used, the successive states of the network are given by,

$$x_j(t+1) = \sum_{i=0}^{n} f_i(x_i(t)) W_{ij}(t). \tag{3.3.9}$$

When neither facilitation nor supression is used, $W_{ij}(t)$ is set to the connection weight between node we and j, that is $W_{ij}(t)=w_{ij}(t)$. If facilitiation exists in the connection, $W_{ij}(t)$ is set to,

$$W_{ij}(t) = w_{ij}(t) \prod_{k \quad F_{ij}} f_k(x_k(t)), \tag{3.3.10}$$

where F_{ij} is the set of indices of the nodes facilitating the connection between node we and j. Similarily, if supression is used,

$$W_{ij}(t) = w_{ij}(t) \prod_{k \quad S_{ij}} (1 - f_k(x_k(t))), \tag{3.3.11}$$

where S_{ij} is the set of indices of the nodes supressing the connection between node we and j.

In all simulations reported below, *Eulers method* is used to calculate the activities of the integrating nodes. It is well known that this method gives rather large errors (Eldén and Wittmeyer-Koch 1987), but since it has worked well in the simulations in this book, we see no reason to use a more complicated method. It also makes it simple to use both integrating and summing nodes concurrently. The formula (3.3.9) is simply replaced by,

$$x_j(t+1) = (1-\varepsilon_j\alpha_j)x_j(t) + \varepsilon_j\sum_{i=0}^{n} f_i(x_i(t))W_{ij}(t). \tag{3.3.12}$$

For a summing node $\varepsilon_j = \alpha_j = 1$, for a leaky integrator $0 < \varepsilon_j << 1$ and $0 < \alpha_j < 1$, and finally for an integrator node $0 < \varepsilon_j << 1$ and $\alpha_j = 0$.

System Properties

When nodes are connected together into networks, it is often interesting to study the long-term behavior of the system. Given that the network receives constant input from the environment, one of three things can happen. In the first case, the activities of the nodes converge to some fixed value. When this happens, the network can be said to have made a choice. The final state is called a fixed point in the state space of the network. In the second case, the network goes through a sequence of states which recur with a fixed time interval. The network is said to have reached a limit-cycle. In the last case, the network exhibits chaotic oscillations. When this happens, the activity of the nodes goes through a never-ending sequence of states that never repeats itself. This last type of behavior is generally undesirable.

In the artificial nervous systems which will be developed in this book, it is usually required that parts of the network reaches either a stable state or a stable limit-cycle. However, this is generally not a desirable property of the nervous system as a whole. A stable state in the nervous system would mean constant behavior throughout the lifetime of the creature. The entirely reactive creatures we will discuss in chapter 4 are exceptions. Since their internal state always reflects the input they receive, they will always be in a stable state when the input is constant. This is not much of a problem, however. Since the creatures are moving around, their input will change all the time, and so will their internal state.

Another question to ask about the internal state is whether it is stable or not. If the state changes slightly, will it return to the previous fixed point or limit-cycle? If it does, the prevous state was stable. Otherwise it was unstable. Since these properties have already been formally studied by many researchers (Amit 1989, Cohen 1983, Kamp 1990), we will not do so here. Instead, we will try to summarize some important architectures which we will use later on in this book.

Oscillation A simple oscillator can be constructed by letting an integrating node with a threshold output function, *f*(*x*), inhibit itself when its activity passes over the threshold (figure 3.3.5). Given a constant input signal, *s*, the activity, *x*, of the node will gradually build up until the threshold is reached. At this point, the node will generate a brief output pulse which will reset the node and the cycle will repeat itself again. For a fixed threshold, the frequency of the emitted signal will be proportional to the intensity of the input signal.

The oscillator is one of the most important types of units used in motor control (Gallistel 1980). Since the creatures in this book will have very limited motoric abilities, the oscillator will not be used to any large extent below.

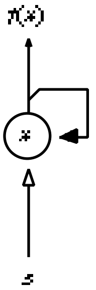


Figure 3.3.5 A simple oscillator. See the text for futher explanation.

Competition One important property of many neural networks is the ability of the nodes to compete with each other and to form a global choice. This process is implemented by letting all the nodes gradually inhibit each other until only one node is at its supra-threshold level. Figure 3.3.6 shows two simple networks with this property. In each network, the three nodes $x_1 \ldots x_3$ receive different input signals called $s_1 \ldots s_3$.

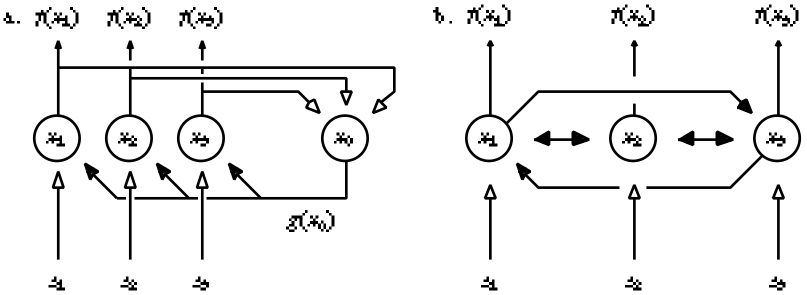


Figure 3.3.6 Two simple competitive networks. All signals except the largest are dynamically quenched by recurrent inhibition.

Let us first consider the network to the left. The output of each integrator node x_i is given by a sigmoid function. Competition is implemented by an auxiliary node called x_0. This node calculates the sum of all the output signals $f(x_i)$ and generates recurrent inhibition through the semi-linear output function $g(x_0)$. This output signal inhibits the nodes $x_1 \dots x_3$ with equal strength. Given that the activities of $x_1 \dots x_3$ build up gradually, the activity of x_0, and consequently also the level of inhibition $g(x_0)$, will gradually increase. When the process stabilizes, only the node with the largest input signal will remain at its supra-threshold level. The output signal of this node, say x_1, will be equal to its input signal s_1. Figure 3.3.7 shows the development of the output signals in a simple competitive situation. As can be seen, all outputs start to increase, but all except for the strongest are quenched by the recurrent inhibition. At equilibrium, the output $g(x_0)$ is equal to the maximum of the input signals. This network can, thus, be used to calculate the maximum of a number of signals.

Since networks of this type have been much studied in the neural network literature (Amari 1982), we will not develop the analysis further here. We will only note that the network to the right in figure 1.4.4 shows almost identical behavior to that to the left. In this network, all nodes inhibit all other nodes proportional to their output signal. This has the same effect as the inhibition from the auxillary node in the network to the left. The only difference is that the outputs are not summed at an extra node. Since the inhibitory connections protrude to the sides of each node, the network is said to use *lateral inhibition*. Inhibition of this type is very common in biological neural networks.

Amari and Arbib (1977) and Amari (1982) analyse networks with recurrent inhibition through an a extra node. A similar network, but based on different principles, is presented in Trehub (1991). Grossberg (1973) has studied the behavior of networks with lateral inhibition combined with different output functions and update rules. He has also shown how choice can be combined with normalization of the input pattern and contrast enhancement in a single network. These properties require quite complex dynamics and will not be used here. We will instead use different, and simpler, network architecture for these processes (See section 7.2 for an example of normalization).

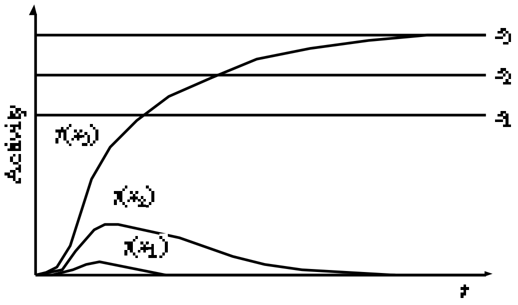


Figure 3.3.7 The development of the different activity levels in the nodes of the networks in figure 3.3.6.

Cooperation A content addressable memory can be constructed by connecting a number of nodes together as shown in figure 3.3.8. This is called a *recurrent* network. Each threshold output signal, $f(x_i)$, is connected back into the network to all other nodes through plastic connections of varying strength. An input signal, s_i, will first activate its corresponding node, x_i. The output from this node will then activate all other nodes that it is connected to. These, in turn, will propagate the activity to the nodes to which they are connected, and so on. This process is called *spreading activation* (Rumelhart and McClelland 1981).

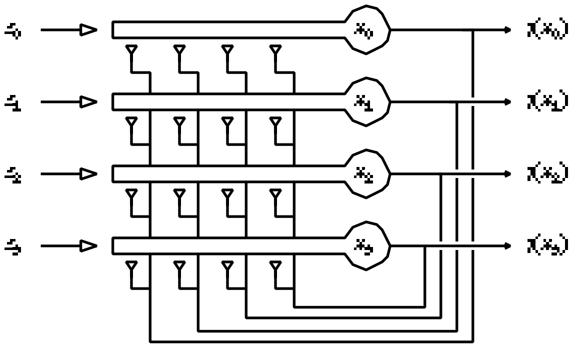


Figure 3.3.8 A content addressable memory.

If two nodes, x_i and x_j, both activate the other in this way, they are said to cooperate. If one node in a set of cooperating nodes is activated, the mutual cooperation will set up a positive feedback loop which will eventually activate all nodes in the set. We call the activity pattern that results a *resonant* state (Grossberg 1989, Shepard 1984). This property can be used to construct a content addressable memory (Grossberg 1989).

Let us call the input to the network $s = <s_0, s_1, s_2, s_3>$ and the output $f = <f(s_0), f(s_1), f(s_2), f(s_3)>$. Let us further assume that we want to store the two patterns $\alpha = <1, 1, 0, 0>$ and $\beta = <0, 0, 1, 1>$ in the network. This is accomplished by setting the connection from node we to j, w_{ij} to 1 if there exists a pattern where both position we and j are 1. To store pattern α, w_{01} and w_{10} are both set to 1 since pattern α contains 1:s at position 0 and 1. Similarly, for pattern β, w_{23} and w_{32} are set to 1. All other connections, including the connection from each node to itself, are set to 0.

Figure 3.3.9 shows the resulting network. If we set the input to any part of the pattern α, say <1, 0, 0, 0>, the output will be the whole pattern, in this case, <1, 1, 0, 0>. The same is true about the pattern β. if the last half of this pattern is given as input, that is, s = <0, 0, 0, 1>, the output will approach <0, 0, 1, 1>. Since each pattern is associated with itself, a system of this kind is sometimes called an *auto-associator* and a content addressable memory is sometimes alternatively called an auto-associative memory.

It is clear that it is easy to store any number of patterns in a memory of this kind as long as they do not overlap. If we try to store two patterns which do overlap, such as <1, 0, 0, 0> and <1, 1, 0, 0>, the largest pattern will always be recalled. This problem gets even more severe, if we try to store a number of patterns that partially overlap, say <1, 1, 0, 0>, <0, 1, 1, 0> and <0, 0, 1, 1>. Giving the network any of these patterns as input will recall the pattern <1, 1, 1, 1>, which was never stored at all.

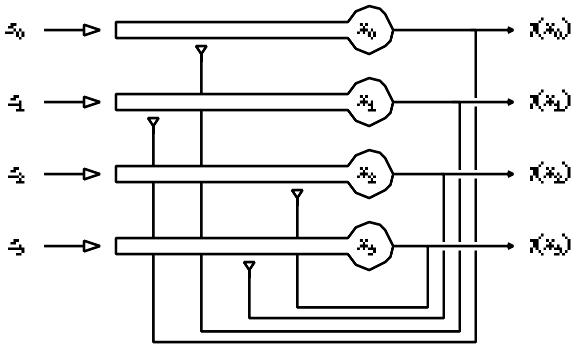


Figure 3.3.9 A content addressable memory which has learned a number of patterns.

A number of methods have been developed to handle these problems. The most obvious one is to let the connections have varying strengths and to include inhibitory connections which make it possible for the different nodes to compete with each other and this increases storage capacity in many cases. Another strategy is to add high-order nodes which recognize specific activity patterns and enhance the storage of these (Cohen and Grossberg 1987). We will return to these problems many times in the discussion of associative learning below. Here we will only present an overview of two different learning strategies which can be used in a network of this type.

There are essentially two methods to set the weights of a recurrent network. The first is to increase the connection weights each time two nodes are active together as in the example above. This method was pioneered by Hebb (1949) and its different variants are often referred to as Hebbian learning rules. The central idea is that temporal *contiguity* of activation increases the connection, or *association*, between two nodes. The second method is to let the connections between nodes represent the statistic *contingency* between them. Networks of this type can be based, for example, on Bayesian decision theory (Lansner and Ekeberg 1989) or on an estimate on the transferred information between the nodes (Balkenius 1992).

The distinction between these two methods for formation of associations has been much discussed whithin animal learning theory (See for example Mackintosh 1983). The important distinction between these two learning types has not been much discussed among neural network researchers.

Conclusion

This section has introduced the basic neural network concepts which we will use throughout this book. A number of node and connection types have been presented. We have also taken a brief look at the phenomena of competition and cooperation in small networks of interacting nodes. In the next section these concepts will be put to work in our first example of a complete artificial creature.

3.4 A Complete Artificial Creature and Its World

The principles presented in the previous section can be incorporated also in very simple creatures. In this section, we want to present one such example which will also serve as an introduction to the methodolgy used in the rest of the book, but first we will describe the world that our creatures will inhabit. Since the creature constructed in this section is mainly intended as an example, we will not put forward much argument in support of the different design choices made here. In fact, some of them are not even very good. In order not to clutter the presentation, such arguments will have to wait until the following chapters.

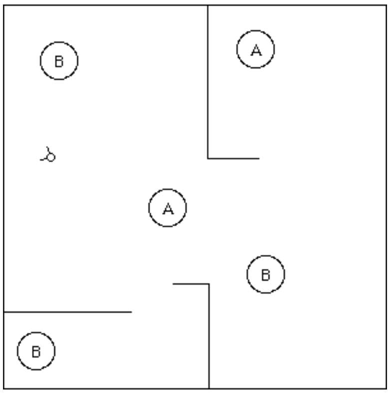


Figure 3.4.1 An example environment. Four walls, three food objects and a single creature are present.

Elements of the Environment

An animal is nothing without its environment. It supplies it with food, shelter and all other things that are essential for a living organism. If we want to study complete autonomous creatures, this can only be done if we also investigate the world which it inhabits. Since the creatures we will consider in this book are all simulated, their world differs in a number of ways from the ‘real’ world. It is, thus, of great importance to fully understand the details of this alternative world before we can start the construction of our artificial creatures.

The simulated environment consists of an infinitely large two dimensional plane, though in practice, the environment is constrained by walls. Creatures and objects are placed in this plane and are all simulated as two dimensional shapes such as circles and squares. In this respect, at least, the environment is not entirely unlike Abbott's Flatland fantasy although the behaviors of its inhabitants are entirely different (Abbott 1884/1991).

There are a number of objects in this world such as walls, food and creatures. It has not been the goal to make the environment as realistic or as complex as possible. While very complex environments can give us insights which are not possible in a simple environment, it is also very often important to strip the environment of all but the essential features. It is useful to consider the world described here not as a model of reality but rather as an alternative world which shares some aspects with the real world. If we were to carry out the task described in this book in the real world with robots instead of simulated creatures, many details of the creatures would have to be different. However, the general methodology would be the same.

Walls Walls are represented as lines in the plane. Walls can be either opaque or non-opaque. Opaque walls do not let odors through while non-opaque walls do. The primary function of walls is to structure the environment into something a little more interesting than an empty surface.

Food There are four types of food in the environment: A, B, C and D. Food is depicted as circles with varying radius and contains a varying amount of energy. They also produce a smell. This smell, that may be a complex composition of odors, can be detected by the creatures and it can guide them towards the food. What type of food it is can be detected by the creature once it is in contact with it and when it eats, a certain amount of energy is transferred from the food object to the creature.

This energy can be of four types called e_0, e_1, e_2 and e_3. Different activities of the creature use different types of energy which makes it necessary for our creature to eat different types of food containing different energy types.

Aversive Objects An aversive object gives a creature an 'electric shock', if it comes into contact with it. These objects are circular or points, just like food, and also give off smells. When a creature is in contact with an aversive object, its shock sensor reacts.

Other Creatures Apart from the creature, which is in our focus of attention, other creatures may also be present in the environment. Usually, these will be copies of our current creature but at times they will be what we will call irrelevant creatures which simply move around at random. The role of these creatures is to annoy the real creature and make sure the environment is unstable.

This list of objects is not complete and we will introduced various other things as we go along, but they are sufficient for our first example of a neurally controlled artificial creature.

The Body of an Artifical Creature

Our creature has a perfectly circular two dimensional body. This may not be much of a body, but since we will mainly study spatial behavior and not motor control, it will be quite sufficient since it contains the most important features of a body, namely sensory receptors and a motor system as well as a simple metabolism. While the nervous sytem of the creature will be developed into considerable complexity through the rest of this book, the body will not be changed at all. By commiting ourselves to a fixed body, we know that a better performance is the result of the developed nervous system and not of a new and better body. The different sensory and motor systems of the creature are described briefly in figure 3.4.2 and below. (See Balkenius 1994a for further details).

Touch The creature has one whisker on each side of the body which is directed forward. The whiskers have their origin at the center of the body and can have a variable length. The angles between the forward direction and the whiskers are always the same for both whiskers. The whiskers have outputs which tell the creature whether or not they have collided with any object such as a wall or another creature. They do not react on contact with food. There is also a sensor in the body which reacts when the creature has collided with an obstacle. This signal can be used as a last resort when the whiskers have failed to react. This is how this sensor is used in the simulations reported below. The neurons which are activated when the whiskers are in contact with an object are call w^L and w^R. The collision sensor is denoted by c.

Smell A number of smell sensors are located at the end of the whiskers. These are called $s^L{}_0 \ldots s^L{}_7$ and $s^R{}_0 \ldots s^R{}_7$ for the sensors on the left and right side respectively. The smell sensors detect the ‘concentration’ of a number of different ‘chemical’ substances in the environment around the creature. The primary sensory system of our creature will be concerned with smell and we will develop the analysis of odors to some complexity in the following chapters.

Taste When the creature is in contact with food, one of the four food detectors will react. There is one food detector for each of the four food types A, B, C and D. These detecors are called f_A, f_B, f_C and f_D. It is possible for the creature to eat a food object only when the food detector reacts on it. Apart from generating eating behavior, the food detectors can be used as signals which initiate learning. The food detectors can be considered to generate taste signals to the creature.

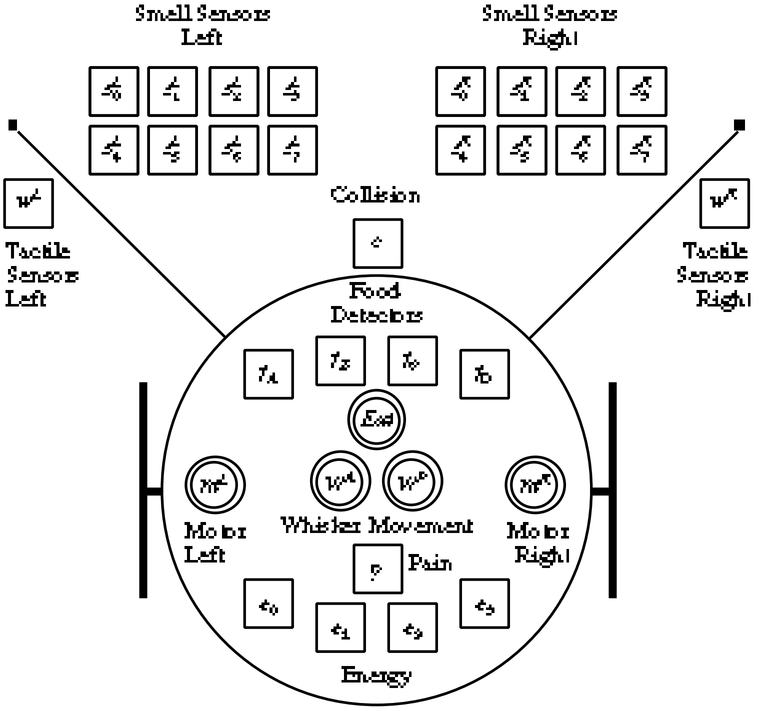


Figure 3.4.2 The body of the creature with its sensory and motor neurons.

Pain The body is also susceptible to ‘electric shock’. Various aversive object generates these signals, if the body is in contact with them. This makes the single pain receptor, *p*, react. Like the food detectors, the pain sensor can be used to drive learning.

Needs Our creature also has sensors for its bodily needs. To keep things simple, all the artificial creature needs is to eat at regular intervals. There are four internal need sensors which determine the internal level of four types of energy needed by the creature called $e_0 \ldots e_3$. Based on the signals from these sensors, the creature can make choices about what to eat.

Note that the outputs from these sensors indicate that the creature does not need the corresponding energy. This means that the creature should look for food containing energy 0 when the sensor e_0 does *not* react.

Motor System The body moves through the environment using two motors which drive wheels on each side of the body. By varying the speeds and directions of the motors, the creature can move forwards or backwards, turn while moving or spin. This type of general body architecture has been used in a number of studies both with physical robots and in simulated environments, most notably by Braitenberg (1984). In these studies, the body and motor control is not the primary object of study. However, the type of output signals necessary to control the movement of the creature can easily be adapted to more developed models of locomotion (for example, Beer 1990, 1992, Ekeberg 1992) or to robots (see, for example, Hirose 1993).

In the simulations shown below, all the movements of the creature are generated by a model of a physical motor system with the only exception that the creature is weightless and the wheels have infinite friction on the ground. The consequence of these simplifications is that the creature has no physical momentum and that the rotations of the wheels are the only thing that move the creature. If the wheels have stopped, the creature is absolutely still. The simulated motor system is a sufficiently good approximation of a robot moving slowly, but not very appropriate if we want to compare it with the precise movements of a fast running animal.

All the movements of the creature are controlled by a set of four motor neurons. Two of these, m^L and m^R, control the speeds of the motors on the left and the right side of the body. Another neuron called *Eat*, controls the eating behavior of the creature. When this neuron is activated, the eating behavior starts. Neurons of this type are often referred to as command neurons and can be found in many lower animals (Shepherd 1988). The last two output neurons, w^ϕ and w^p, control the angle between the two whiskers and their protrusion respectively. By changing the output level of these neurons, the creature is able to move its whiskers (see section 4.4). The different motor neurons are shown in figure 3.4.2.

Metabolism Our creature has a simple metabolism. The energy gained from eating decreases at a rate which depends on the current activity of the creature. The faster a creature moves, the larger the decrease will be. This metabolic model makes it favorable for a creature to rest when it is not in need of food since this will lower its metabolic rate.

If we want to evaluate the performance of a creature, the type of metabolism we select will play an important role. While it is tempting to evaluate a creature based on how good it is at solving a problem, such as finding a hidden food pellet, this can never be our only criterion. Real creatures spend a lot of energy while searching for food and if we do not take this into account, the best creature would be one that generates very strange behavior.

For example, if the creature is rewarded for finding and eating a lot of food but not penalized for moving fast, it need not be very smart. The best strategy will be to run as fast as possible to cover the whole of the environment at a minimum of time. If, on the other hand, it is expensive to run, it is much more beneficial for the creature to use a more clever strategy. This implies that a performance measure must be based both on the observable behavior of the creature *and* on the amount of energy used to carry out the task.

A Complete Creature

Since we are primarily interested in cognitive abilities, the parts of the creature described above will be held fixed during the development of the various successively more complex nervous systems. In this way, we will know that any altered performance of the creature is a result of its new nervous system and not of some changed property of its body or sensors.

We are now in a position to construct a simple creature which conforms with the principles described in the previous section. The first principle tells us that we should start with a description of the behaviors that we want the creature to use, starting from the most essential and working towards the more complex. The final creature will consists of a set of distinguishable subsystems which we will introduce one at a time. When the creature is complete, we will investigate the various evolutionary sequences which could have constructed that creature.

Let us first construct a complete artificial creature which simply wanders around and explores its environment at random. This example is similar in many respects to the artificial cockroach (*Periplaneta computatrix*) described by Beer (1990).

Subsystem A: Move A simple explorative behavior consists of two parts, one behavior which moves the creature around and a second which lets it avoid obstacles. Moving around is accomplished by simply giving the motors constant input signals. If the input signals given to the two motors are different, the motor speed will differ and the creature will move in a circle. This is useful since it will make the creature cover a larger portion of the environment than movement in a straight line.

Figure 3.4.3 shows the motor nodes necessary to generate this behavior. The left motor is controlled by the node m^L and the right motor by the node m^R. The resting levels of the two nodes are shown below each node.

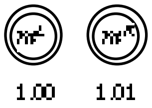


Figure 3.4.3 The ‘network’ that generates the basic move behavior. Only the two output nodes are needed. Since the resting activity of the left node (1.00) is slightly lower than that of the right (1.01), the creature will move in a large circle.

Figure 3.4.4 shows the path taken by a computer simulated creature with such a behavior. As we can see, it almost immediately get stuck at a wall. Fortunately, the motors do not run at exactly the same speed and the creature will slowly turn and finally get away. This is possible only since the simulation shown used what we may call ‘slippery walls’ (see Balkenius 1994a). Such walls let the creature turn even though the force generated by its wheels is in the direction of the wall.

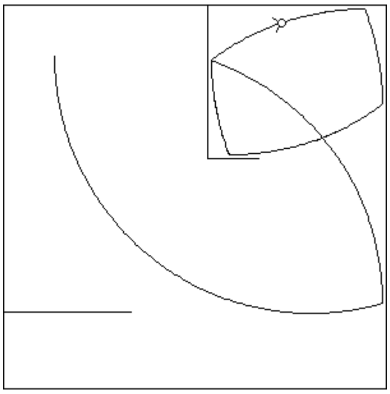


Figure 3.4.4 The route taken by our first computer simulated creature. It will stay in the same path forever.

Subsystem B: Avoid There are two problems with the behavior of our current creature. The first is that it uses most of its time at walls slowly turning, and the second is that it very easily gets stuck in a loop as can be seen in figure 3.4.4. We avoid the first problem by making use of the whiskers. When the left whisker is in contact with a wall, we let the creature increase the speed of the left motor. The right whisker will increase the speed of the right motor in a similar way. The extra signal to the motors will make the creature turn away and successfully avoid obstacles much faster than the previous creature.

This is a simple example of a subsumption hierarchy since the turning behavior overrides the moving behavior. Combining behaviors in this way will be one of the main subjects of chapter 4. There is one more turning behavior to consider, however, and that is the situation when the creature has walked straight into a wall and both whiskers react. Since the speed of both motors will increase, the creature will continue to move straight ahead into the wall.

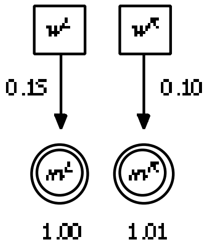


Figure 3.4.5 The controlling network for Move and Avoid. The input nodes for the whiskers on the left and on the right are called w^L and w^R. The motor nodes are called m^L and m^R.

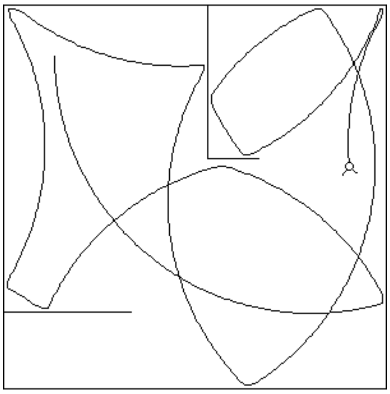


Figure 3.4.6 The path taken by the creature with Move and Avoid behavior.

This situation can be avoided if the connections from the whiskers to the motors have different strengths. In this case, the motors will have different speeds and the creature will slowly turn until only one of the whiskers reacts. When this happens, the creature will turn faster and eventually get away from the wall.

The necessary controller is shown in figure 3.4.5. The touch signals from the two whiskers enter the network at the input nodes w^L and w^R. When any of these are active, its output signal is sent to the motor nodes m^L and m^R respectively. The numbers at the connections indicate the connection strengths on the connections from the touch sensors to the motor nodes. The behavior generated by this controller is shown in figure 3.4.6.

It is interesting to note that although the control system can be considered behavior-based, there is no strict correspondence between modules in the network and overt behaviors. The same connections in the neural network are used to let the creature turn left or right and to avoid straight-ahead collisions. The combined abilities can be considered a behavior module for different types of avoidance. The network in figure 3.4.5 can be described by the subsumption hierarchy shown in figure 3.4.7.

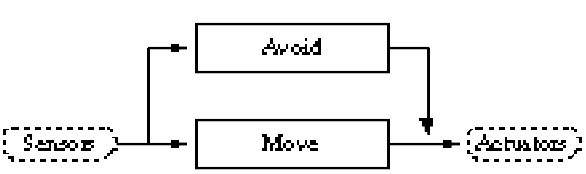


Figure 3.4.7 The subsumption hierarchy corresponding to the network in figure 3.4.5.

Subsystem C: Explore The path taken by our new creature is much better than the first example since it does not get stuck in loops in the simple example environment. In more complex environments this is still a problem however and something more is needed. In Beer (1990) it was suggested that a good exploratory behavior could be constructed by using a very simple strategy. The creature turns slowly to the left or to the right while moving and changes between these two behaviors after random intervals. There are two reasons for using such a behavior. The first is that the randomness of the walk makes sure that the creature will not get stuck and the second is that the large circling movement will eventually cover the whole environment. If we want our creature to explore, this is obviously a good thing.

To make the creature able to avoid walls, we keep the previous architecture and simply add a few neurons that make it change between the turning right and turning left behaviors. The network used here is different from the one described in Beer (1990), but his solution would probability be possible here, too.

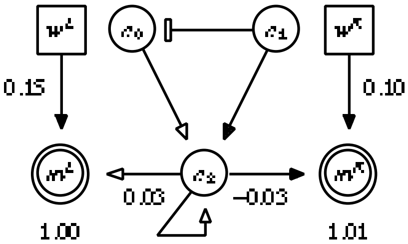


Figure 3.4.8 The neural controller for the exploratory behavior *Move+Avoid+Explore*. The random circular walk is controlled by a stochastic sample-and-hold circuit.

To change between the two behaviors, a random sample-and-hold curcuit was constructed (figure 3.4.8). A threshold summing node with feedback, n_2, is used to store the current mode of the creature. The state of this node can be either 1 or 0. If this node is active, the creature will turn left and if it is inactive, the creature will turn right. This is our first example of a neural controller with an internal state. However, since the state is set at random, the behavior of the creature is not disrupted if it loses its state for some reason.

To set the turning mode at random, two noisy neurons were used to change between the two states. The first noisy neuron, n_0, is active or inactive with equal probablity while a second neuron, n_1, is allowed to burst with a small probability. Every time n_1 generates an output signal, it first resets the summing neuron, n_2, and then samples the output of n_0 by facilitating it. If n_0 is active, the mode neuron, n_2, will be set again and otherwise it will stay reset. Since the output of n_0 has equal probability for both states, the probability for each of the two modes is the same.

Figure 3.4.9 shows the behavior generated by this controller. As can be seen, it does not only cover a large area of the environment and prevents the creature from getting stuck in loops, it also shows some occasional wall following behavior which is a combination of the turning and obstacle avoidance behaviors. Wall following can readily be considered as an emergent property of the network controller. There is no special wall following module in the nervous system, but the creature will engage in this behavior anyway. Below, we will develop a set of more complex exploratory behaviors.

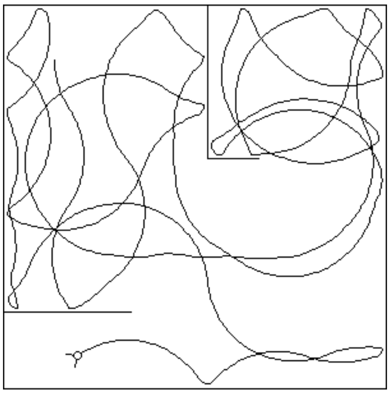


Figure 3.4.9 The exploratory behavior generated by the nervous system in figure 3.4.8.

Subsystem D: Eat Walking around exploring the environment is of no use in itself, of course. The only reason for our simple creature to do so is that it will be able to find food which is not accessible from its initial location. Eating behavior can easily be added to our creature since it is already equipped with both taste sensors and an eating command node. We let the creature eat as soon as its taste sensors react. We need to add very little to the controller to make this possible. We connect each of the sensors for pallative food with the eating command node as shown in figure 3.4.10. This network works even if it is totally disconnected from the movement controller described above. This would be a simple example of completely distributed control.

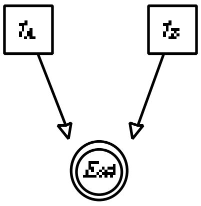


Figure 3.4.10 Each of the food detectors for pallative food starts eating behavior.

Subsystem E: Stop at Food A better behavior is generated, if the eating system is allowed to inhibit the exploratory behavior to let our creature slow down when it tastes food. This will make it possible to eat more than if eating has to be done, so to speak, on the fly. This admittedly simple view of eating is perhaps sufficient for a very simple creature, but we will see in chapter 6 that much more complexity is needed to generate a good eating behavior.

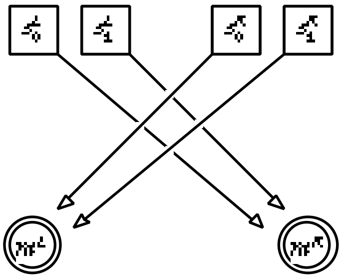


Figure 3.4.11 A goal directed network. The network guides the creature toward objects that smell of smell 0 or smell 1.

Subsystem F: Goal Direction Our current creature will have great troubles finding food since its movement is not at all goal directed in the sense that it specifically moves the creature towards food. Such a behavior can be constructed by using smell cues in the environment. Since the two smell receptors are placed apart from each other on each side of the body, the differences between the left and right smell can be used to guide the creature towards food. Figure 3.4.11 shows a network based on this idea.

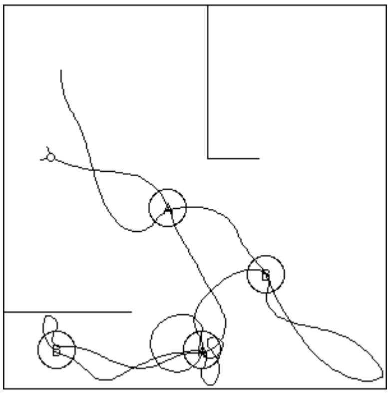


Figure 3.4.12 Goal-directed behavior. The creature turns towards food when it is sufficiently close to smell it.

When the smell of food is more intensive on the left than on the right, the right motor will increase its speed more than the right one and the creature will turn left towards the food. This design was suggested by Braitenberg (1984) as an example of the simplest way in which goal direction could be constructed. There exists a number of variations on this simple circuit and we will consider them at length in chapter 4.

The creature with this goal directing system uses an approach–consummation strategy. When it smells food, it will turn towards it and increase its speed until it reaches the food where it will slow down and start to eat. When all the food is eaten or when it has left the food patch, it will continue to explore the environment (figure 3.4.12).

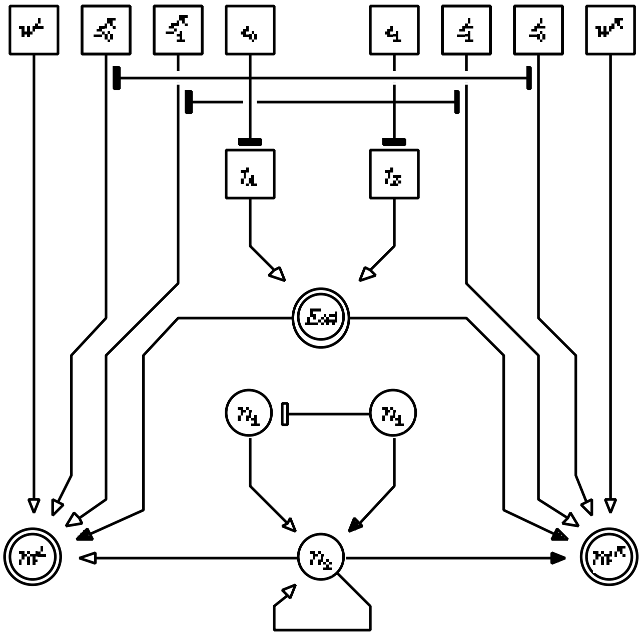


Figure 3.4.13 A complete creature which uses many of the mechanisms described in the text. Note that the smell sensors are connected to the motors on the opposite side of the creature.

Subsystem G: Central Engagement Selection There is one problem with our current creature. It will never stop eating if there is food around since it has no way of knowing if it needs more food or not. A way to avoid this problem is to let the creature monitor whether it needs more energy or not. Figure 3.4.13 shows an example of such an architecture. Two need sensors, e_0 and e_1, are added which react when the creature has no need for a certain food type. These sensors inhibit the food detectors as well as their corresponding smell receptors. As a result, the creature is insensitive to the types of food it does not currently need and will consequently ignore them. This is a simple form of motivationally biased perception. In chapter 7, this type of mechanism will be the basis for a discussion of attention.

The system consisting of the two need detectors and their inhibitory influences on the rest of the controller is a simple system for central behavior selection. We see that this system only shows one of the three aspects of such a system, namely internal drive. The introduction of internal and external incentives will be postponed until chapter 6, where their relation to motivation will be investigated.

Conclusion

This section described a complete creature showing how the design principles introduced in section 3.2 could be incorporated in a very simple creature. Our example creature implemented the principles of behavior-based control, subsumption, parallel engagements, and central engagement selection. Most of all, it is entirely reactive. In the next chapter, we will take a closer look at different types of reactive behavior. In the later chapters of the book, we will take a closer look at learning, motivation and perception in general, and in our artificial creature in particular.

Chapter 4

Reactive Behavior

4.1 Introduction

Behavior in its most elementary form is purely reactive. The motor patterns produced are caused only by the immediate perceptual state of the animal. In this chapter, we will introduce some basic classes of reactive behaviors and consider what different properties they possess. There are two distinct historical lines that we will follow.

The first historical line has its start in Lewin's topological psychology (Lewin 1936) and his view of behavior as controlled by a force field generated by objects or situations with positive or negative values or *valences*. The idea was later developed by Hull (1932, 1943, 1952), who used it as the basis for his goal-gradient construction. After having fallen in disregard for many years, the idea has been recently resurrected through Arbib's (1987, 1991) work on motor schemata and Arkin's (1987, 1990) use of potential field methods for robot navigation. The common denominator of all these approaches is their account for the *directedness* of behavior. They all subscribe to the position of *hedonism*[1], that is, the organism is considered to approach pleasure and avoid pain. This leads naturally to a distinc-

---

[1] In ethics, the doctrine of hedonism argues that humans should strive for pleasure and try to avoid pain. The corresponding position in psychology argues that this is already the case. Note also that the term is inherently circular. It appears to be impossible to define pleasure and pain without making any reference to the fact that people will strive for the first and try to avoid the second.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

tion between four types of behaviors: (1) *appetitive* behavior that is directed toward an object or a situation, (2) *aversive* behavior that is directed away from an object or a situation, (3) *exploratory* behavior that relates to objects or situations that are unknown, and finally, (4) *neutral* behavior that is guided by objects that are known to be neutral, that is, neither appetitive nor aversive.

The other historical line is that of reflexes and fixed motor patterns. The idea of reflexes has a long history, but the word was not used until 1784 when it was introduced by Prochaska (See Shepherd 1988). The concept of fixed motor patterns was suggested by Lorenz (1950). By introducing *hierarchies* of motor patterns, Tinbergen (1951/1989) made one of the most important contributions to this area. Recently, the development of behavior-based robots moves the idea in a new direction, from a descriptive framework to a foundation for design (See, for example, Brooks 1991a, and Maes 1990). While the first historical line is mainly concerned with directedness, this other line emphasizes *coordination* and *sequences* of motions.

The two directions are not at all incompatible, of course. In fact, one of the conclusions of the present chapter will be that a better understanding of reactive behavior must be based on directedness as well as coordination and sequences of behavior. The chapter concludes with an inventory of basic reactive behaviors.

4.2 The Directedness of Behavior

One of the most fundamental aspects of most behaviors is that they are directed. Animals do not just move around at random but are guided by their perception of the world in such a way that their behavior is likely to accomplish something of importance, such as finding food or a mate. It will thus be necessary to consider goal-direction at some length before we can start the construction of an artificial creature. To make the notion of direction precise, we will need to define a language suitable for the task. This language will take its starting point in Lewin's topological psychology. We will tailor it to fit our needs, and as the language becomes more precise, it will gradually begin to look like a theory of behavior.

To make the presentation clearer, we will start with behaviors that are purely reactive and relate to objects in space. This also makes it natural to use spatial concepts to describe behavior. As the theory moves away from reactive behavior, the spatial concepts will gradually be transferred from the external space to the internal space of the animal. In the later chapters, we will try to make a case for the idea that all cognition is based on spatial concepts and that goal-directed thought may merely be goal-directed behavior made internal. Before we can make such broad claims, however, we must first move back in time to the early days of topological psychology.

Topology and Vectors: Lewin's Field Theory

The psychology of 1930 left much to hope for in terms of descriptive precision. It is not hard to understand why the psychologists of the time envied the exactness and conceptual rigor of physics and mathematics. Inspired by the latest developments in pure mathematics, Kurt Lewin tried to developed a formal mathematical language for psychology. The foundations for this language were presented in his book *Principles of Topological Psychology* (Lewin 1936), which was the first in a series of works that developed his ideas from a topological psychology into a *vector psychology* where the behavior of a person was explained in terms of *forces* acting on the individual.

The problem that hampered the development of a formal vector psychology was that it seemed difficult or impossible to define a metric on the psychological space. Without a metric, the concept of psychological vector or force is meaningless. In this respect, vector psychology could never offer more than a metaphorical description of a person. However, as we will see below, for artificial creatures, at least, a precise vector psychology is in fact possible. To make such a vector theory possible, we must first consider Lewin's theory in more detail.

The psychological theory developed in the language offered by topology and vectors is generally referred to as the *field theory*. The central concept of the field theory is that of a *life space* that determines the behavior of an individual. The behavior at any moment in time can be described as a function of the life space of the individual. The life space includes the whole situation determining the behavior of an individual, that is, both the person, P, and the environment, E. If we denote the behavior of the individual by B, the field theory can be summarized in Lewin's famous formula,

$$B = f(P, E). \tag{4.2.1}$$

One of Lewin most important insights is that we should consider the *momentary* situation of an individual in order to understand its behavior. This contrasted with many of the popular theories at the time, where the causes of behavior were sought not in the present but in the past[2]. This does not mean that the past does not have a role in determining the present situation of an individual. On the contrary, both the experienced past and the expected future, that is, the total life situation, may play a part in determining the immediate situation. But, it is this situation that determines behavior and not the past or the future.

> The goal as a psychological fact undoubtedly lies in the present. It really exists at the moment and makes up an essential part of the momentary life space. On the other hand, the content of the goal … lies as a physical or

---

[2] Strangely enough, this is also true of modern behaviorism that allows any time interval between stimulus and response.

> social fact in the future. Indeed, it may not occur at all. The nature of the expectation and the character of what is expected, in so far as they act as psychological conditions at the moment, naturally do not depend upon whether or not the event comes to pass. (Lewin 1936, p. 36)

Although Lewin did not consider this possibility, the language that he developed makes it possible to talk about all behavior as reactive in a sense. However, since reactions have their causes in the momentary life situation, and not only in the physical environment, the complexity of behavior can go far beyond simple stimulus–response reflexes.

It is in the representation of the psychological environment that Lewin's mathematical concepts reveal their large potential. Like von Uexküll's *Umwelt*, the psychological environment refers to the environment as it is perceived by the individual and not to the physical environment (See section 2.9). Consider, the following example from Lewin (Lewin 1936 p. 96; figure 4.2.1). A boy wants to become a physician. To achieve this goal, he is required to first pass the college entrance examination (ce), then college (c), and medical school (m), etc. Finally, he will be able to establish a practice (pr) and reach his goal (G).

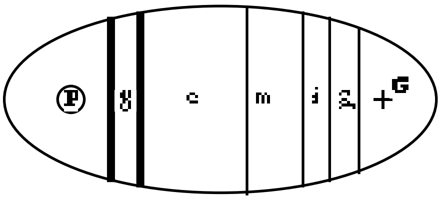


Figure 4.2.1 The situation of a boy that wants to become a physician. P, person; G, goal; ce: college entrance examinations; c, college; m, medical school; i, internship; pr, establishing a practice (adapted from Lewi, 1936).

Figure 4.2.1 illustrates the main components of topological psychology. The situation consists of a number of *regions*. One of them contains the person, P, and another is the *goal* region, G. As can be expected, regions can be included in each other. In the figure, the different regions are included in the total situation characterized by the outer oval.

Regions are separated by *boundaries*. These are considered to resist *psychological locomotion* to a degree that can vary from nearly none to infinity. Boundaries that offer considerable resistance are called *barriers*. In this example, psychological locomotion refers, not to bodily motion, but to an alteration in position in the quasi-social[3] life space of the boy. In the boy's conception, the college entrance

---

[3] Three important parts of the life space are the quasi-physical, quasi-social and quasi-conceptual facts. These are the internal representations of the external reality that influences the behavior of the person.

examinations offer the greatest resistance, represented by thicker lines in the figure. This is probably, as Lewin noted, a false but nevertheless clear picture.

In this example, the track to becoming a physician is represented by a set of spatial regions. Although the artificial creatures we will develop below will not have such high ambitions, it is still possible to use a similar language to describe their behavior since the problems they come across will usually be spatial in a very concrete sense, as well. For example, to obtain food, the creature must move from one location to another. This makes the two-dimensional domain that was introduced in the last chapter ideal for the exploration of topological concepts.

The concept of a life space can readily be used to represent the momentary situation of our creatures as well as the influences on behavior exercised by the past and the expected future. We will see in the next chapter that learning can be seen as a process that restructures the life space in such a way that the resulting behavior is changed. In the more general case, it is also possible to consider concept formation and problem solving as processes of restructuring.

So far, we have deliberately avoided Lewin's vector concepts since we will have reason to discuss them at length below. However, a few simple examples are appropriate here. Figure 4.2.2a shows a child (C) on one side of a barrier (B) and a desired toy (T) on the other. The force field acting on the child is indicated by an arrow.[4] It is obvious that movement in this direction will not reach the desired goal since the barrier is in the way. Figure 4.2.2b shows a restructured life space that would solve the problem. The dotted path shows one of the possible ways around the barrier. In figure 4.2.2c the child finds himself in a situation which contains an aversive object that causes the child to withdraw or retreat.

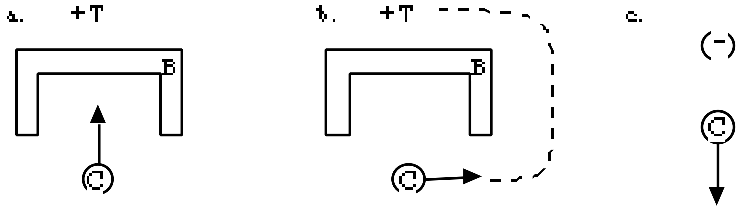


Figure 4.2.2 Field forces influencing the behavior or a child (C). (a) The desired toy (T) directs the locomotion of the child in its direction. (b) A restructured life space that makes it possible for the child to reach the toy. (c) An aversive object (–) makes the child withdraw.

---

[4] In Lewin's studies, vectors are drawn as 'pushing' the person and not 'dragging' as here. This, however, is a misrepresentation of the vector concept that caused Lewin to make some predictions that at times did not strictly follow from his theory.

We see that forces can be either *driving*, such as forces generated by the desired or the aversive objects, or *restraining*, such as the force corresponding to the barrier (Lewin 1935). In general, the closer to the generating object or event the person gets, the larger the driving forces become. We will return to the different types of field forces that can be represented in this way below where we outline a modern version of the vector psychology.

When using ideas from physics psychologists have been known to make misstakes from time to time and Lewin is no exceptions. One of them is to confuse the notions of velocity and acceleration. This leads to strange conceptions such as force-fields controlling speed rather than acceleration. This is unfortunate since it makes it impossible to use the terminology of psychology without being in conflict with the use of these terms in physics. Since the terminology has already been established in both fields, some abuse of terminology is thus unavoidable.

Hull's Goal-Gradient Hypothesis

The idea of force fields as the determinants of behavior was also incorporated in Hull's theory in the form of goal-gradients (Hull 1932, 1938). Although stated in somewhat different languages, both Lewin's and Hull's theories make essentially similar prediction in many situations. This is especially true in their analysis of conflict and choice. However, Hull's theory moved beyond Lewin's when it tried to give an explicit account of the process that constructs the goal-gradients. It also tried to explain how the goal-gradient is transformed into movement in the physical world. This is something that Lewin's theory could not explain.

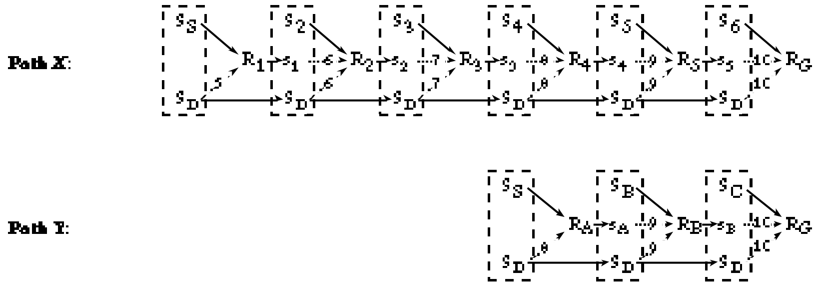


Figure 4.2.3 The stimulus–response dynamics of a simple maze. Behavior is explained with the following ideas: drive stimulus, S_D, stimuli in the environment, S_i, proprioception, s_i, reactions, R_i. S_S represent the stimuli in the start box and R_G the goal reaction. (Adapted from Hull 1932).

Figure 4.2.3 shows the stimulus–response dynamics of a simple maze learning task when the animal has learned it correctly. Suppose there are two possible paths from the start box to the goal box, one long, X, and one short, Y. A process that conditions reactions to the stimuli preceding it sets up a goal-gradient for the environment. All reactions, R_i, are conditioned according to the temporal and spatial proximity to the goal reaction R_G. The reaction that directly preceded the goal reaction is most strongly conditioned to its preceding stimuli and the earlier reactions are conditioned progressively weaker.

The upper series of stimuli, S_i, represent external stimuli that can be detected in various places in the maze. The S_D represents an internal drive stimulus, such as hunger, that persists throughout locomotion in the maze. The small s's represent proprioceptive[5] stimulation that results from the execution of their preceding responses. The dotted arrows represent learned associations set up during the learning phase and the small figures represent the strengths of these associations. As can be seen, the closer an association is to the goal reaction, the stronger it is. Also note that the two sequences have the same start S_S.

When the animal is placed in the start box and is under the control of the drive stimulus S_D, two reactions are available to it that have previously been reinforced, R_1 and R_A Since the reaction R_A is closer to the goal reaction, both in time and space, the *reactive potential* is larger for R_A than for R_1. As a consequence, the animal will chose the reaction R_A that will result in the shortest path to the goal, that is, path Y.

The reactive potential of an animal is considered to increase with the proximity to the desired goal, thus making the animal run faster the closer it comes to it. The original formulation of the goal-gradient hypothesis made many unrealistic simplifications. It ignored that an animal has a top speed and that the goal-gradient can, therefore, not be directly reflected in the speed of movement. It is also necessary for the animal to slow down before reaching the goal in order not to collide with it.

Despite these simplifications, Hull developed the force-field concept in a very important way when he made the medium representing the forces explicit. This medium was the set of responses available to the animal. Using the set of responses as a starting point, it is possible to define both the direction and the magnitude of the forces determining behavior. For such a definition to work, however, it is necessary that responses are defined in relation to the animal, as is the case in Hull's theory. This is not possible within a theory that defines responses in relation to the environment as, in Skinner's operants, for example (Skinner 1974).

The responses available to the animal determine the directions of the forces influencing its behavior in a self-centered coordinate system. The reactive potential of these responses gives the magnitude of the forces controlling the animal. Given

---

[5] Proprioception refers to all sensory input from the musculoskeletal system. By using this concept, Hull's theory did not have to rely on any lasting internal state of the organism. Fortunately, the contemporary philosophical climate does not require us to make such a perverse intererpretation of Hull's theory. There is no reason why we could not interpret the $R_i \rightarrow s_i$ association as an internal representation of the previously performed action.

a certain stimulus situation and a complete description of the animal, we can potentially calculate the forces acting on the animal and predict its behavior.

The responses are similar to the boundaries in Lewin's topological psychology, but instead of viewing them as resistance, we may identify them with a certain amount of effort. Thus, the resistance of a boundary is equal to the effort required to execute the reaction that moves the animal through the boundary. Using this odd but important identification, the situation in figure 4.2.3 can be redrawn in a Lewinian flavor as shown in figure 4.2.4. This representation of the situation could be enhanced by including vectors that would represent the strength of the force field, that is, the same information as the reactive potentials in the Hullian analysis.

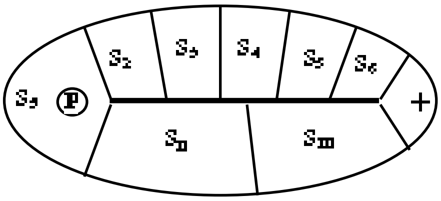


Figure 4.2.4 The dynamics of the simple maze using a Lewinian representation. Each stimulus is contained in a region of the life space. To pass from one region to the next, an animal must cross a boundary, that it, it must perform some action. The boundary in the middle is assumed to resist any attempt to cross it. Thus, there are only two paths from the start to the goal.

Potential Fields and Motor Schemata

Both Hull and Lewin used a physical language to talk about psychological phenomena. This is even clearer in Köhler's version of the field theory. He used the analogy[6] of electromagnetic force fields to describe the dynamics of behavior. This approach is highly holistic since it emphasizes the role of the whole field as a determinant of behavior. A single stimulus can, thus, not be considered as the cause of behavior until it has been incorporated in the whole.

In the use of potential fields methods for robot control, the parallelism to physics is made even more explicit. The robot is considered to be an electrically charged particle in an electric field generated by the various objects in the environment. Obstacles are assigned negative potentials while the goal is considered positive. If the robot follows the gradient of the electric field, it will successfully reach the goal and avoid obstacles in most environments.

---

[6] Sometimes he meant it literally, though. The electrical activity recorded on the skull was considered as a sign of these internal electric fields.

The method was introduced by Khatib (1985), but is primarily known through Arkin's (1987) work in the area. He has combined the potential field method with the concept of a motor schema introduced by Arbib (1991, 1993). A motor schema is a representation of a certain motor activity with respect to every location in the environment. Figure 4.2.5b shows the motor schema for the action *move-to-goal* of the environment in figure 4.2.5a. The arrows shows that the direction of movement at each location in space is in the direction of the goal point.

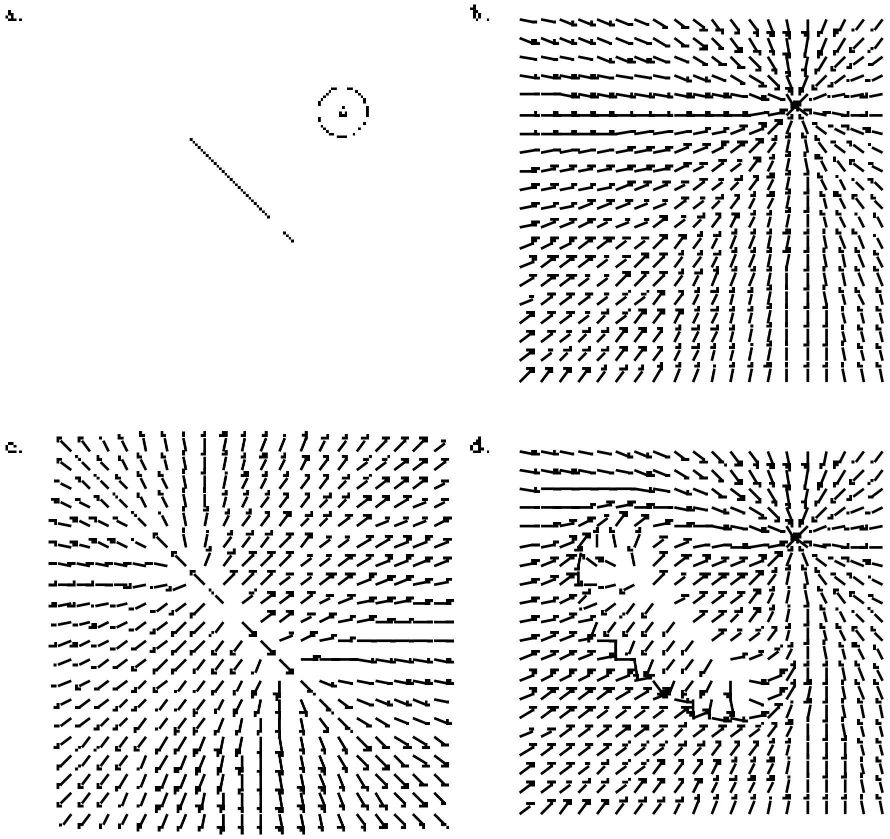


Figure 4.2.5 Two motor schemata and their combination. (a) The environment. (b) The motor schema *move-to-goal*, (c) The *avoid–wall* schema, (d) The behavior of a creature using the combined schema.

This particular motor schema could have been generated by the potential field method since the direction of movement coincides with the attracting force of the electrically charged particle at the goal point. For an environment with no obstacles, this motor schema is sufficient to navigate toward the goal. In a more complex environment, more schemata are needed. Figure 4.2.5c shows the motor schema *avoid-wall* considered as a field generated by the wall. The robot is supposed to move away from the wall.

If the wall and the goal are located in the same environment, the action of the robot can be calculated by adding the two force fields together as shown in figure 4.2.5d. Note that the picture shows the path preferd by the robot rather than the sum of the two vector fields. By following this combined motor schema, the robot can reach the goal without colliding with the obstacle. The most important aspect of this control method is that it generates stable control strategies. It does not matter if the robot is accidentally moved off course since the gradient field will automatically compensate for this and get it on the right track again. Another nice property of the method is that the different motor schemata controlling the robot can be calculated independently and then simply combined, either by adding them or by taking the maximium of them. For instance, the processes that calculate the *move-to-goal* schema do not communicate with the *avoid–obstacle* process at all. The potential field method is, thus, very well suited for parallelization (See also section 4.3).

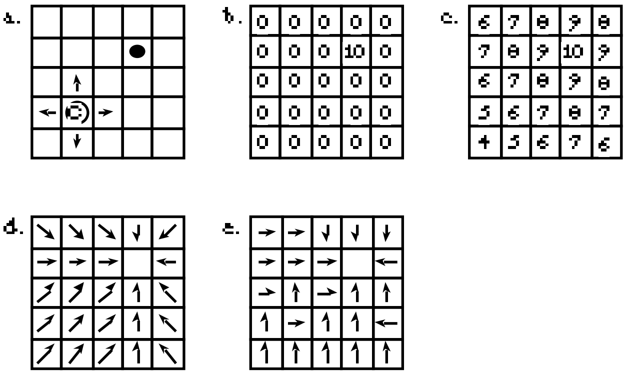
Although this method has many merits, there are a number of important problems; the most noticeable being that it very easily generates local maxima in the gradient landscape at other places than at the goal. One possible solution to this problem is to add noise to the motor schemata, but this is not entirely satisfactory (Arkin 1990). Another problem is that it is location based. This means that, in general, the location of the robot must be known before the force vectors can be calculated. Potential fields may also give rise to strange behaviors, such as avoidance of objects behind the robot (Connel 1990).

Despite these problems, the idea of motor schemata is very attractive as a way to represent the latent behavior of an agent. In many respects, it is similar to Lewin’s vector fields, but instead of showing the direction of movement in an internal psychological space, it shows the direction of movement in the physical space. If we keep the state of the agent constant, we can, potentially, move the agent to all possible locations in space and record its movement. This record is essentially an image of the psychological environment of the agent and we have, thus, found a way to make the notion of a psychological space precise. This is one of the central ideas that we will use in the next section as we try to outline a modern field theory.

Toward a Modern Field Theory

Though in theory, one should be able to measure all the parameters controlling an animal in a given situation, this is not possible in practice. With artificial creatures, however, the situation is different. The problem here is not to keep track of all the parameters, but rather to make sense of them. The modern version of the field theory that we will outline in this section is intended to describe very complex behavioral dispositions and internal representations in a way that relates directly to the behavior they may potentially cause.

Like Lewin's topological psychology, the concepts put forward below are initially intended as a language and not a theory in itself. They are, however, highly interwoven with the theory we will develop. We also believe that this language will aid the presentation in the rest of the book since it allows a common description of many different perceptual and behavioral phenomena. The language will be approached by a number of progressively more complex examples.



0	0	0	0	0
0	0	0	10	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

6	7	8	9	8
7	8	9	10	9
6	7	8	9	8
5	6	7	8	7
4	5	6	7	6

Figure 4.2.6 A discrete environment. See the text for an explanation.

Let us first consider the discrete environment shown in figure 4.2.6a. It consists of 25 distinct regions. One of them contains the creature, C, and another contains the goal, marked by a black dot in the figure. The arrows show the four possible directions of movement the creature is able to perform. Like in Lewin's topological psychology, the boundaries between the different regions offer a resistance to locomotion. Let us assume that the strength of this resistance is 1. This means that the *effort*, or *cost*, required to pass through the boundary is also 1. We let the function $c(a, b)$ denote the minimal cost of moving from location a to b. Note that this cost function does not need to commute as it does in the example, that is, it is not necessary that $c(a, b) = c(b, a)$.

The reward obtained at each region is shown in figure 4.2.6b. As we can see, the creature receives a zero reward at all locations except for the goal. The values shown in this figure consitute the reward, or goal, function for the environment.

Figure 4.2.6c shows how much it is worth to be in each of the different regions. These values are called the potential reward, G, of each location, that is, the reward the animal will obtain if it follows the path that offers the least resistance to the goal. If $R(g)$ is the reward obtained at the goal location, g, and z is the current location of the creature, the potential reward function is defined as,

$$G(z) = R(g) - c(z, g). \tag{4.2.2}$$

This suggests a definition of *psychological distance* between two regions as the psychological, or perceived, effort required to move from one region to the other using the path with least resistance. For psychological distance between a and b, we will use the notation, $\Psi(a,b)$. This measure will be formally defined in chapter 5. Note that psychological distance does not in general define a metric on the life space since the psychological distance from A to B may not be the same as that from B to A. This non-commutativity of psychological distance is a well known fact from psychological experiments (See, for example, Lee 1970).

If we want an artificial creature to behave optimally, it should be designed in such a way that its psychological distance coincides with the actual cost function of the environment. In the next chapter, we will consider a set of learning mechanisms that have as their function to establish approximations of appropriate measures of psychological distance.

Figure 4.2.6d shows the direction of maximal growth of the potential reward. To receive the maximal total reward the creature should follow these directions as closely as possible, that is, choose the path with the shortest psychological distance. So far we have only discussed what the creature should do. To see what it really does, we need to observe its behavior in the environment.

Figure 4.2.6e shows one possible outcome of placing the creature in each region and recording its action. Since diagonal locomotion is not possible, the creature has to choose randomly in some regions. Note that we do not need to make any assumptions about the internal workings of the creature to make a record like this. However, it does require that we can somehow keep the internal state of the creature constant. For an artificial creature, this simply means that we temporarily turn off any learning abilities it might have.

Given a result like that in figure 4.2.6e, what can we say about the creature? We may consider the actions selected as an approximate record of the life space of the animal. To get a better picture, we could place the creature in each region many times and calculate its average direction of movement. Let us assume we received the set of directions shown in figure 4.2.6d, that is, on the average, the creature followed the direction of maximal growth of the potential reward. From this, we would conclude that the goal is represented in the life space of the creature. We also see that the environmental representation of the animal must agree, to some extent, with that in figure 4.2.6c.

This conclusion, however, requires that we first establish the base case for the behavior of the creature. In the following, we will assume that its behavior is random when no objects are perceivable in the environment. Random behavior will be marked with a black dot.

Let us now introduce a barrier in the environment as shown in figure 4.2.7a. Figure 4.2.7b, c and d show three possible outcomes if we repeat the procedure above in the new environment. The different records make us draw different conclusions about the life space of the creature. In the first case, 4.2.7b, the creature obviously knows the location both of the barrier and the goal and can choose the optimal path. What the field does *not* tell us, however, is *how* the creature knows the optimal path. The historical cause of the behavior can be excessive training in the environment. However, it is also possible that the creature can directly perceive both the barrier and the goal and can make the sufficient inferences about the optimal path. It may even have some form of innate knowledge of the world.

In the situation in 4.2.7c, the barrier hides the goal which makes the creature move at random at places where the barrier is in the way of the goal. In figure 4.2.7d, the goal is perceived by the creature but the barrier is not. Its movement is identical to that in the previous example. The result of this representation is that the creature will get stuck at the barrier if it starts on the wrong side of it. If it selects actions at random, it may get away eventually, however.

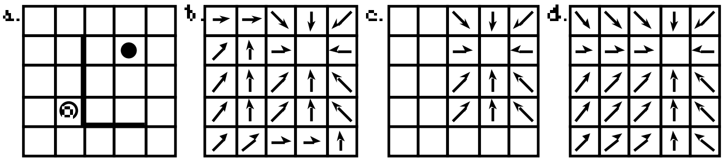


Figure 4.2.7 A discrete environment with a barrier.

This example shows that it is possible to represent perception in terms of action. In figure 4.2.7b, where the creature acts on both the barrier and the goal, we may say that it has somehow perceived them both. In figure, 4.2.7d, on the other hand, the creature has perceived the goal but not the barrier.

This view of perception emphasizes its active nature and its relation to action. *Something is perceived if it can be acted on*, that is, if it changes the life space of the animal. Remember that all the arrows in each of the figures above do not represent actual movement, but rather potential movement that will be executed only when the creature is placed in that particular region of space. Perception is, thus, not identical to action, but rather to latent action (See chapter 7).

In the examples above, we have made many simplifications, the most important being the use of a discrete set of regions and actions. To make the ideas presented above more realistic, we must make the transition to continuous spaces. For this, we first need to define some mathematical concepts. Given a fixed coordinate system, with points $x = (x_0, x_1, \quad , x_{n-1})$, a scalar field is described by a function, say, *f*, that assigns a scalar value to every location in space. From a differentiable scalar field we may derive a *gradient field*, *f*, that gives the *direction* of maximal growth of the scalar field for all locations in the space. The operator is pronounced ‘del’ or ‘nabla’, and in *n* dimensions it is defined as:

$$\equiv \left(\frac{\partial}{\partial x_0}, \frac{\partial}{\partial x_1}, \quad , \frac{\partial}{\partial x_{n-1}} \right) \tag{4.2.3}$$

A gradient field is, thus, a special type of *vector field*, that is, a function that assigns a vector to each point in space. In physics, gradient and vector fields are used to describe force fields of the type that Lewin used to describe the life space of a person. It should, thus, come as no surprise that the concept of vector fields will be of great importance throughout this book.

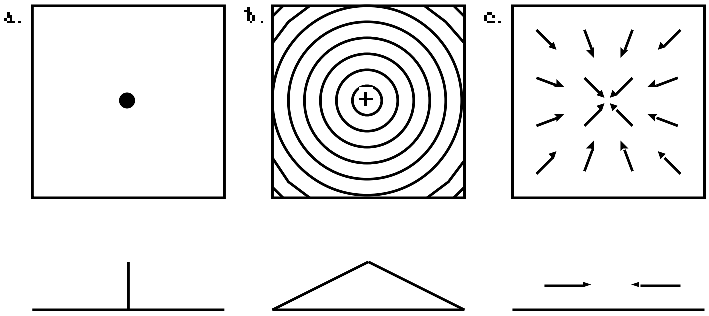


Figure 4.2.8 The relation between the reward function, the potential reward function and the gradient. (a) The reward (or goal) function. (b) The potential reward function. (c) The gradient generated by the potential reward function. The lower figures show cross sections of the three functions.

We can use these mathematical concepts to develop a field theory for continuous spaces in the following way. A goal, *g*, is placed at the location marked by *a dot* in figure 4.2.8a. When the animal reaches this location, it receives a scalar reward, R_g. As above, the environment can be described with a *reward function* R(z), that assigns a reward to every location in the environment *z*. In this simple example, $R(z) = R_g$, if z=g and $R(z) = 0$ otherwise. When reward is received only at the goal locations, the reward function will sometimes be referred to as the *goal function*. From the reward function, we may derive a second function describing the *potential reward* at each location in space. At the goal point, the potential reward is equal to the actual reward R_g. At other places the potential reward is given by the reward at the goal minus the minimal *effort*, or *cost*, required to move from that location to the goal. The cost of moving from *a* to *b* will be called $c(a, b)$. Figure 4.2.8b shows a potential reward function generated around the single goal in 4.2.8a. The optimal direction of locomotion is given by the gradient field derived from the potential reward function, G as shown in figure 4.2.8c. The relation between cost and the potential reward function, G for a goal g, is thus given by the following formula,

$$G(z) = R_g - c(z, g) \tag{4.2.4}$$

The functions depicted in figure 4.2.8 are defined in terms of the environment. It is obvious that the environment must first be perceived by the animal before it can influence behavior. To give a picture of the life space of an animal, the vector fields must be defined in terms of the animal and not in terms of the environment. Keeping this in mind, we can now formally define our equivalent of Lewin's formula.

Let us assume that the location and posture of an animal can be described by a set of values $z = (z_0, z_1, \ldots, z_{n-1})$. Any action that the animal can perform can be represented as a change in z. The *momentary behavior* of the animal is thus given by the quantity,

$$\frac{dz(t)}{dt} = \left(\frac{\partial z_0(t)}{\partial t}, \frac{\partial z_1(t)}{\partial t}, \ldots, \frac{\partial z_{n-1}(t)}{\partial t} \right) \tag{4.2.5}$$

To calculate this quantity, we need a number of interrelated concepts. The *momentary sensory environment*, $s(t)$, picked up by the sensory apparatus of the animal at time t, is a function of both the physical environment, $e(t)$, and the current location in space, $z(t)$. The immediate sensory environment should be contrasted with the *potential sensory environment* given by the set of all sensory signals that could potentially reach the animal if all possible positions where assumed. The potential sensory environment is, thus, a concept related both to von Uexküll's *Umwelt* and Gibson's ambient optical array (Gibson 1979).

The *internal state* of the animal is described by the vector, $x = (x_0, x_1, \ldots, x_{m-1})$. This state changes as a function of time and the momentary sensory environment. When the internal state of the creature is independent of time and is only a function of the momentary sensory environment, the creature is said to be *reactive*.

This definition of a reactive creature is an idealization, since it assumes that there is no time lag between sensory input and actuator output. In reality, no animal can be reactive in this sense, but the definition is, nevertheless, useful since it allows us to discuss reactive behavior in a stringent way. It would, of course, be possible to define the momentary sensory environment as a function of an interval of time instead, but this would make the mathematical presentation unnecessarily complex.

The internal state is used to calculate the *momentary actuator output*, a, of the animal. For a purely reactive creature, the actuator output is given directly by its location in space. Note that the momentary actuator output is different from the momentary behavior. The actuator describes what the creature *tries* to do, and the behavior is what it can be observed to do when the actuator output has been transformed by the mechanics of the body and the laws of physics. This lets us keep the ordinary meaning of the word behavior.

The momentary behavior is calculated from the momentary actuator output, the location in space, and the properties of the environment, together with the physical laws governing that environment. Thus,

$$\frac{dz(t)}{dt} = f\big(a(t), z(t), e(t)\big) \qquad (4.2.6)$$

Figure 4.2.9 shows the different components of the creature and its environment and their relations. Note that the location and posture of the animal must be considered a property of the environment since its alteration depends on the actuator output as well as the environment.

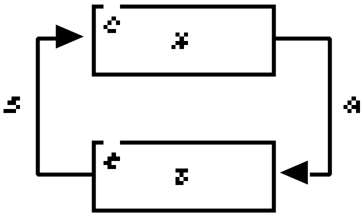


Figure 4.2.9 The interaction of the creature, *c*, and its environment, *e*. An actuator output changes the location of the creature, *z*, which in turn changes the momentary sensory input, *s*. The sensory input changes the internal state, *x*, of the creature which in turn produces a new actuator output. The internal state of the creature and the environment may, of course, also change on their own regardless of each other.

This equation will be our version of Lewin's *B*=*f*(*P*, *E*). There is one important difference however. In our formulation, the concept of a life space is not merely metaphorical. Given that we know the internal state of the animal, it is possible to draw a picture of its life space. Especially in the case of a purely reactive creature, this can easily be done in a way analogous to the discrete situation presented above. For a fixed time, *t*, we estimate the average vector field and take this as the life space of the creature. When the behavior of the creature is entirely deterministic, the life space is given directly as a function of *z*:

$$f_t(z) = f\big(a(x(s(z))), z, e(t)\big) \qquad (4.2.7)$$

This suggests that we view the creature and its environment as a dynamical system (Kiss 1991, Thelen and Smith 1994). In the rest of this chapter, we will take a closer look at this function and characterize some different classes of reactive behaviors in terms of the life spaces they generate.

Behavior in Relation to a Single Object

We now put the definitions above to work and investigate the basic classes of directed behavior and the minimal artificial nervous systems that can generate them. Throughout the presentation, we will use the nervous system in figure 4.2.10 as an example. This network can generate all of the behaviors described in the text. (Figure 4.2.11 summarizes the appropriate connection weights for the different behaviors.) Behavior in relation to a single object in space can be characterized by whether the object is appetitive, aversive, neutral or unknown, which gives us the four classes of behavior that will be presented below.

In the animal learning literature, the different types of behavior listed in table 4.2.11 are usually defined in terms of learning (See, for example, Gray 1975). Here, we will try to make definitions that are entirely independent of any learning abilities that the animal may have and are thus applicable in many more situations.

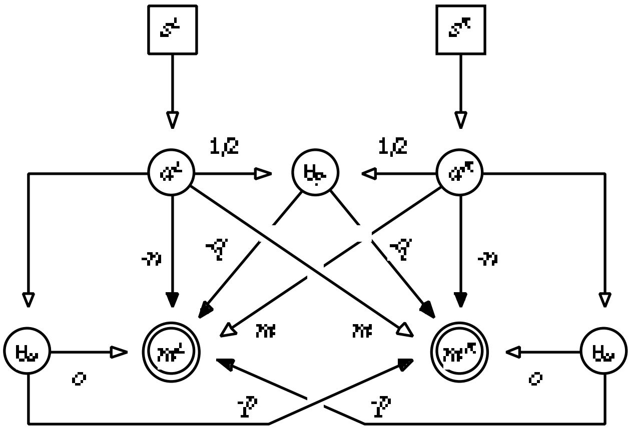


Figure 4.2.10 Nervous system that can generate the different behaviors discussed in the text. Olfactory sensors at the top linearly excite or inhibit the motor neurons at the bottom to produce the desired behavior. Inter-neurons with semi-linear output functions make the avoidance gradients steeper than the approach gradient. (See text for explanations).

Behavior in Relation to an Appetitive Object In order to eat, an animal is usually required to move away from its present location and toward the location of food. Once the food source is reached, the food-seeking behavior is followed by eating. This is an example of the two main components of behavior in relation to an appetitive object. In many cases, both *appetence* and *consummation* may consist of a whole set of behavior patterns with varying degrees of complexity and adaptability (Lorenz 1977, p. 57; Timberlake 1983).

Behavior Type				Connection Weights					
				m	n	o	p	q	r
Approach		Accelerating		a					
		Decelerating			a				c
		Combined	Accelerating	a		b			
			Decelerating		a		b		c
			BIS	a				b	
Avoidance	Active	Accelerating				a			
		Decelerating					a		c
		Combined	Accelerating		b	a			
			Decelerating	b			a		c
	Passive							a	c
Orienting	Toward			a	a				
	Away					a	a		

Figure 4.2.11 Summary of the different behaviors discussed in the text. The right part of the table shows connection weights that should be used for the nervous system in figure 4.2.10 to produce the various behaviors. The values *a* and *b* are arbitrary constants with $0<b<a$ and $c>0$ is the resting activity of the motor neurons. Blank table entries indicate that the corresponding connections are missing or have zero weights.

Here we will consider the dynamic properties of the appetitive situation as they appear in the spatial domain when only one goal object is present. In the next chapter, we will study how different learning methods are involved in the appetence and the consummation phase, and in chapter 6, the role of time varying valences will be investigated.

Appetitive behavior can be divided into the two classes of *goal-directed* and *non goal-directed* or *wandering*. To distinguish the two, we will present a definition of goal-directed behavior and consider behavior that does not fit this definition as non goal-directed.

We want to propose that the following definition captures the essential properties of goal-directed behavior. Let *g* be the location of a goal, *G*. A behavior, *f*, is *goal-directed*, with respect to a stationary goal *G*, if it has the following properties:

(GD1) The location of the goal influences the behavior, that is, *f* is a function of *g*.

(GD2) The distance to the goal in psychological space decreases with time, that is,

$$\frac{d}{dt}\psi(z(t), g) < 0.$$

Given this definition, it is neither necessary nor sufficient that the animal reach the goal. This definition of goal–directedness is, thus, primarily concerned with the striving of the behavior and not with its result. Also note that goal–directedness in this sense is a momentary property of a behavior. Since goal–directed behavior can alternatively be characterized as approaching the goal, this type of behavior can also be referred to as *approach behavior*. This does not mean that all behaviors which approach a goal necessarily are goal–directed. It is quite possible for an animal to approach a goal without utilizing an approach behavior.

The definition should be compared with the one presented by McFarland and Bösser (1993) which requires that the goal is explicitly represented and is used to guide behavior over time. The differences between their notion of goal-directedness and the one we present here are very subtle, but it appears that their definition is slightly more restrictive. Condition (GD1) appears to be weaker than the requirement for explicit representations and condition (GD2) does not require that the difference between the desired state of affairs and the current state is explicitly used as feedback to control behavior, as they suggest. In the ethological literature, approach behaviors, as defined above, are usually referred to as *topic responses* or *taxes*.

> Common to all these taxes, the simplest and the most complex alike, is the fact that the creature turns directly, without any process of trial and error, in the direction most favorable for its survival. In other words, the size and the angle through which it turns is directly dependent on that formed by the creature’s longitudinal axis and the direction of the impinging stimulus. This ‘angle-controlled’ turning is characteristic of all taxes. (Lorenz 1977 p. 52)

The mechanisms that control taxes are called *servo-mechanisms* (Gallistel 1980). They are, thus, canonical examples of teleological behavior (Rosenbleut, Wiener and Bigelow 1943). To model goal-directed behaviors we will elaborate on the idea of potential reward functions. We have already seen how a potential reward function can be defined around a goal. To describe the dynamics of the appetitive situation, some further functions of a similar kind are needed. We will call these *goal-gradients*[7] in agreement with Hull’s use of the term (Hull 1932). The goal-gradient around an appetitive object describes the strength of the approach tendency at every distance from it. The size of the goal-gradient is considered to be reflected in the speed of movement at different distances. The concept of a goal-gradient is, thus, related to the potential reward function described above.

In our artificial creatures, the smell gradients picked up by the olfactory sensors generate signals that have the desired properties of a goal-gradient. Taking this

[7] Here, the word gradient refers to a value that is graded by the distance to the goal. The meaning is thus different from that used above in relation to gradient fields. This use of the word is unfortunate but the two different meanings of the word are already established within psychology and mathematics respectively.

into consideration we see that there are four possible types of goal-directed approach behaviors (See figures 4.2.10 and 4.2.11). The first will be called *accelerating*. A behavior of this type becomes more vigorous as the psychological distance to the goal decreases. The opposite behavior is also possible and will be called *decelerating*. A behavior of this kind will typically cause the animal to stop at the goal. The behaviors of real animals usually show a combination of these two types. Up to a certain distance from the goal, the speed of movement increases until it reaches its top speed. After this point, the speed decreases instead and nears zero at the goal. This will be called a *combined* approach behavior. Note that a combined approach behavior can be generated by combining either an accelerating approach and an accelerating avoidance behavior or by combining two decelerating behaviors.

Miller (1944) suggested that the combined approach behavior is the result of desired objects being both appetitive and aversive, thus generating both an approach gradient and an avoidance gradient. This corresponds to the idea that a goal object may be dangerous if the animal collides with it at high speed. Since the slope of the avoidance gradient is considered to be steeper than that of the approach gradient, the sum of the two gradients will yield an approach behavior that increases in speed as the animal gets closer to be goal but then drops off to zero when the goal is reached. Schmajuk and Blair (1993) have shown that if these ideas are combined with an appropriate physical model, the speed of movement generated closely follows that of a real animal in a runway.

Finally, a creature can show *orienting* behavior. Such a behavior moves the animal closer to the goal in the sense that it directs its sensory apparatus toward the goal. Table 4.2.11 shows the different types of goal-directed behavior, and in figure 4.2.10 a nervous system that can generate the different behaviors is shown.

Figure 4.2.12 shows computer simulations of the three types of approach behaviors using the type of smell cues described in the previous chapter. Accelerating approach causes the creature to move toward the goal. When the goal is reached, the locomotion of the creature starts to oscillate around the goal point (4.2.12a). Decelerating approach makes the creature move more slowly the closer it gets to the goal. In the simulation in figure 4.2.12b, the creature stops before it reaches the goal. Finally, in figure 4.2.12c, a combined approach behavior is simulated. The resulting path shows an abrupt change at the distance from the goal where the approach and the avoidance gradient are equally strong. At this point, the creature starts a circular locomotion around the goal point. The distance from the goal to this circle depends on the relative strengts of the approach and the avoidance gradients. While the nervous systems simulated here were explicitly constructed, behavior similar to that shown in the figure has also been automatically evolved using simulated evolution (Beer 1992). Very similar architectures have also been discussed by Braitenberg (1984).

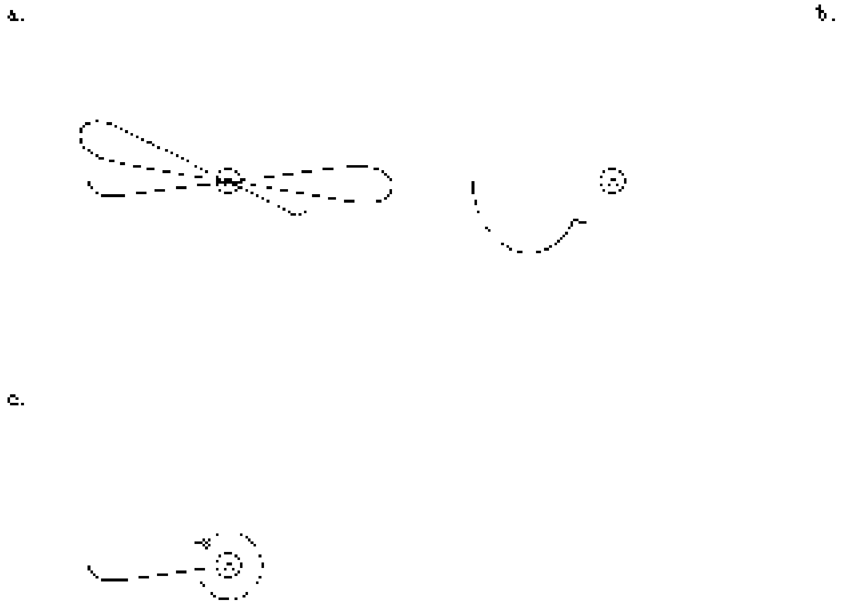


Figure 4.2.12 Computer simulations of four classes of approach behaviors. (a) Accelerating approach followed by oscillations around the goal. (b) Decelerating approach: the creature stops before it reaches the goal. (c) Combined approach: the creature is attracted to a circular path around the goal. Note that in all these simulations, the creature does not stop when it reaches the goal.

In the combined approach behavior, the approach tendency is inhibited by what will be called an active avoidance tendency that increases as the animal gets closer to the goal. On its own, this active avoidance tendency would make the animal move away from the goal, as we will discuss further below in relation to aversively motivated behavior. Animal studies suggest that there also exists another avoidance tendency, usually called passive avoidance (Gray 1975, Mowrer 1960/1973). The role of this avoidance gradient is solely to inhibit another behavior, in this case the approach behavior. To do this, it should not be directed. Thus, it can not be used to move away from an object. In figure 4.2.10, a central node, θ_p, is included that allows behavior to be inhibited in this way. Following the suggestion of Gray (1975, 1982), this system will be called a behavioral inhibition system (or BIS), and the behavior it produces will be called passive avoidance. This is a system of great importance that we will return to many times below. Here, we will only consider its role in slowing down the creature before it reaches the goal.

A creature using a passive avoidance gradient in a combined approach strategy must have a correct estimation of the distance to the goal if it wants to approach it successfully. If the estimated distance to the goal is too short, it will stop before the goal is reached, and if it is to far, the creature will either run passed the goal or get hurt while colliding with it. To behave optimally, the creature should stop exactly at the goal; that is, the approach and avoidance gradients should be equal at the goal location.

In figures 4.2.10 and 4.2.13, the connection weight m represents the slope of the approach gradient, and n is the slope of the passive avoidance gradient. The intersection of the avoidance gradient with the abscissa is given by the threshold, θ_p, of the interneuron whose transfer function is assumed to be semi-linear. The creature will stop when the positive and negative goal-gradients are equal. At this location, g, the smell intensity, s, is given by,

$$s(g) = \frac{n\theta}{n - m} \tag{4.2.8}$$

For optimal speed, n, m and θ should be chosen to satisfy this condition. Even so, we still need to make one change in the approach behavior to make the speed optimal.

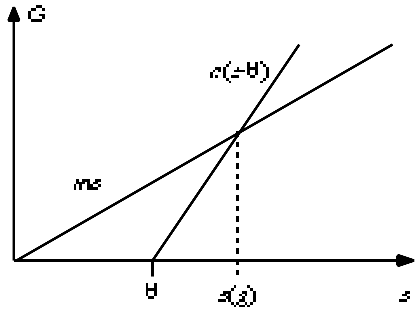


Figure 4.2.13 The approach and avoidance gradients around a goal.

If $G(z)$ is the potential reward function around G, and G is stationary, we see that a behavior is *optimal*, with respect to G, when $f(z) = \ G(z)$. None of the approach behaviors described so far are optimal with respect to the potential reward function of the situation. As we have discussed above, the creature should try to follow the gradient of this function if it wants to reach the goal with the least possible effort. This corresponds to movement along a straight line from the start to the goal, but the creature can only move in this way when it is already heading in the correct direction. An optimal behavior could thus be constructed if the creature could be made to first orient itself toward the goal and then start its approach behavior.

If we let behavior be controlled by both an accelerating and a decelerating approach tendency at the same time, the creature will easily orient itself toward the goal (figure 4.2.10 and table 4.2.11). Since the two motor neurons will always receive equally large signals but of opposite sign, we know that the creature will not move toward or away from the goal but will only turn on the spot. If the left and right sensory signals are different, the creature will turn until these are equal and the goal is in front of it. As we have already seen, this means that the orienting behavior is an approach behavior since the psychological distance to the goal decreases, allthough the physical need not do so.

We also need a mechanism that can change between the orienting and the approach behavior. Figure 4.2.14a shows a hierarchy of behavior modules that can do this. In figure 4.2.14b, the corresponding neural network is shown and a simulation can be found in figure 4.2.15.

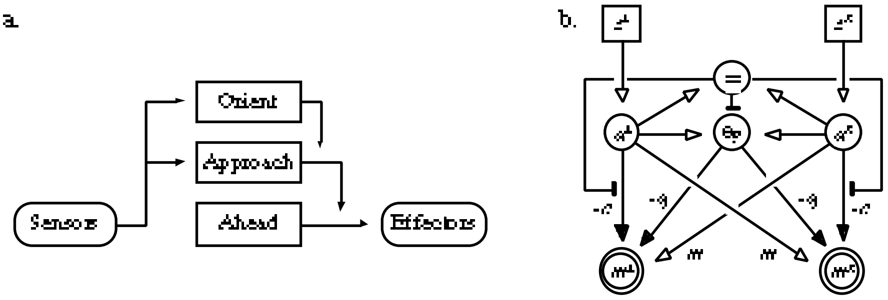


Figure 4.2.14 Combining orientation and approach for optimal behavior. (a) A subsumption hierarchy with three interacting behavior modules. The lower module makes the creature move ahead, the approach module moves the creature toward the goal object and the orienting module orients the creature toward the goal if it is not sufficiently straight ahead. (b) A nervous system that generates the combined orientation and approach behavior. A comparator neuron (=) detects if the olfactory signals from the two receptors are approximately the same. In this case, the inhibitory synapses representing the decelerating approach gradient are inhibited. This will make the creature approach the goal. When the two sensors give different signals, the orienting behavior is disinhibited and the creature will start to orient toward the goal. Connection weights are set as illustrated in table 4.2.11.

The behavior of the creature in figure 4.2.14 is similar to that of a flatworm which orients itself toward a current that carries the scent of food. When both tips of its head receive equal stimulation, it begins to crawl upstream (Lorenz 1977). The nervous system in figure 4.2.14 is optimal for approach to a goal object if it is the only one in an environment. In more complex situations it is not sufficient. In fact, it is not even very good since the creature will orient itself toward the average location of all goal objects in the environment. This is exactly what the flatworm does in such a situation. In section 4.4, which deals with orienting behavior, we will return to this problem and suggest a solution. However, it should already be clear that we have constructed a neural circuitry that can be used as a base for *stimulus–approach* association.

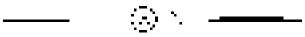


Figure 4.2.15 Orienting behavior followed by an accelerating approach. As can be seen, this type of behavior is more effective than the approach behaviors in figure 4.2.12.

The combined approach behavior described above is our first example of an *attractor* in the life space of the artificial creature. An attractor is a set of points in the life space that has the following properties

(A1) Once the creature has reached the attractor, it will not leave it again.

(A2) Given that the creature is sufficiently close to the attractor, it will eventually end up in it.

Attractors will play an important role as we further develop the field theory. An obvious application is to characterize a goal. When a goal is an attractor for the locomotion of the creature, we know that the goal will be reached if the creature starts out sufficiently close to it. The set of points for which attraction is sure is called the *basin of attraction* for the attractor. This suggests the definition of *goal-reaching* behavior. Such a behavior is generated when,

(GR1) The goal is an attractor, and,

(GR2) the creature is located in the basin of attracton of the goal.

We see that for goal-reaching behavior, the attractor coincides with the goal. Note that the definition of goal-reaching behavior depends on both the psychological environment of the creature and its physical environment. Like the goal-reaching behavior, orienting behavior also ends up in a point attractor. This is an attractor of another type, however, since it is independent of the location variables of the creature. For our example creature, only the posture but not location is influenced by the orienting behavior. This means that for the orienting behavior, the attractor is of lower dimensionality than the action space z; that is, regardless of where the creature is located, it will turn toward the goal stimulus, but will not move away from its current location.

Behavior in Relation to an Aversive Object To survive, an animal must avoid situations that are dangerous. In general, there are two classes of aversive situations that must be avoided. The first type, including events such as jumping down from a high tree or eating poisonous food, does not usually require that the animal performs any specific action. To avoid these situations, all the animal has to do is to refrain from doing something. This is *passive avoidance* that we discussed in the previous section.

The other type of aversive situation requires that the animal take some action; for example, when it is confronted with a predator or a hostile member of its own species. In these cases, the appropriate action is often to escape. This kind of behavior is also called *active avoidance*. Like approach behavior, active avoidance can be performed with different intensity at different distances from the aversive object. For example, it may be adaptive to start out an escape slowly to avoid the attention of a predator and run fast only at a certain distance from the predator. In other cases, it may be more adaptive to run fast when the predator is close but slow down when the distance is larger. A usual strategy is to freeze to avoid the attention of the predator when the distance to it is great and only run when the pretador gets closer.

The different types of active avoidance behaviors parallel the approach behaviors presented above. The difference is, of course, that behavior is directed away from the aversive object and not toward it. Let h be the location of a stationary aversive object, H. A behavior, f, consists of *active avoidance*, with respect to H, if it has the following properties:

(AA1) The location of the aversive object influences the behavior; that is, f is a function of h.

(AA2) The distance to the aversive object in psychological space increases; that is,

$$\frac{d}{dt}\psi(z(t), h) > 0.$$

Condition (AA2) suggests that direction of *optimal avoidance* is given by the gradient of the potential reward function around H. Thus, active avoidance is optimal if $f(z) = \ H(z)$, where $H(z)$ is the potential reward function that is typically negative with its minimum at h. In analogy with approach behavior, it is easy to construct creatures that show the different types of active avoidance behaviors (See figure 4.2.10). Two computer simulations of accelerating and decelerating active avoidance behavior are shown in figure 4.2.16.

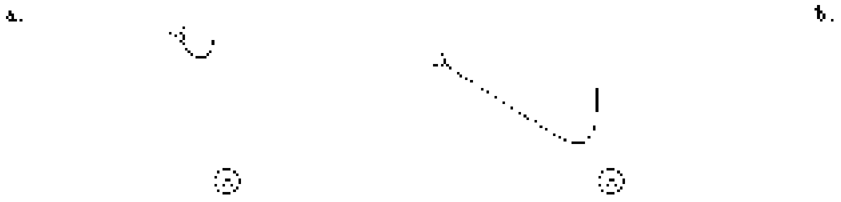


Figure 4.2.16 Computer simulations of two types of active avoidance behaviors. (a) Accelerating active avoidance. (b) Decelerating active avoidance.

Behavior in Relation to a Neutral Object Most elements of an environment are neither very useful nor very dangerous for an animal. Such elements will be called *neutral objects*. In our example world, the only neutral objects are walls and objects that give off neutral smells, but in the real word most objects are of this type. Of course, this does not mean that neutral objects cannot become attractive or aversive at another point in time (See chapter 5). In this section, we will consider objects that are not interesting in themselves at the moment and whose only function is to be in the way of the animal. Such objects are similar to the barriers in Lewin’s field theory in that they hinder psychological and physical movement. However, they are also different in that they may take on aversive properties in certain situations.

Let us first consider the case of simple collision avoidance. In this type of passive avoidance, the animal needs to move around an object in order to reach a goal object. A situation of this type has already been shown in figure 4.2.2. Had it not been for the obstacle, the optimal path would be in a straight line from the current location of the animal to the goal. If we assume that the animal is sophisticated enough to perceive both the goal and the barrier at a distance, and that the influence of the goal is such that the animal is not able to move around the barrier, how should the speed of movement be changed near the barrier if the creature is to avoid colliding with it at high speed?

To avoid collision, the animal must obviously slow down before it reaches the barrier. It seems appropriate to assume that the creature should try to move as close to the goal as possible. It should therefore move the whole way up to the barrier before it stops. This means that its speed must be zero at the barrier, and thus, the speed of movement must be changed in such way that it is positive everywhere along the path but decreases to zero exactly at the barrier.

Since the speed of movement depends on the attractiveness of the goal, the force generated by the barrier cannot be independent of the goal object. The most natural way to make the opposing force of the barrier dependent on the goal is to consider it a function of the speed of movement of the creature, or alternatively of the approach gradient. Let $g(z)$ be the behavior of the creature when the barrier is removed, and let b be the point of intersection between the path to the goal and the barrier. A behavior, f, is called *neutral avoidance* if it can be written on the form,

$$f(z) = g(z) - n\big(\Psi(z,b)\big)g(z) \, , \tag{4.2.9}$$

where $\Psi\big(z,b\big)$ represents the psychological distance between z and b, and $n(x)$ is a scalar function with the following properties:

$$0 \le n(x) \le 1 \, , \tag{4.2.10}$$

$$n(0) = 1 \, , \tag{4.2.11}$$

$$\frac{dn(x)}{dx} < 0 \, . \tag{4.2.12}$$

The expression $n(\Psi(z,b)g(z)$ sets up a *neutral avoidance gradient*. In Hull’s analysis of the above situation, he concluded that the barrier should be considered as an aversive object with an avoidance gradient of a practically vertical degree of steepness (Hull 1952, Theorem 90). In light of the previous presentations, it seems that his conclusion is invalid for animals that do not have negligible mass. Since the animal must slow down before it reaches the barrier, the normal passive avoidance depends both on the speed of movement and the mass of the animal, that is, on its momentum. This cannot sensibly be considered as a property of the barrier. The same problem faces a potential field account of the situation.

Figure 4.2.17 shows the simplest nervous system that is able to stop the creature in the desired way. When the creature comes into contact with a wall, its speed is reduced in such a way that it will stop just before the wall. Since the whiskers of our creature are relatively short, the avoidance behavior will not start until the creature is fairly close to the obstacle. This type of avoidance is, thus, more effective if some other type of sensors are used that react at a longer distance.

The previous discussion of neutral avoidance assumes that there is only one barrier in the environment. Since this is not usually the case, we need some way to combine avoidance gradients from different barriers. Above, we saw that two ways to combine gradients are to sum them together or to take the maximum of them.

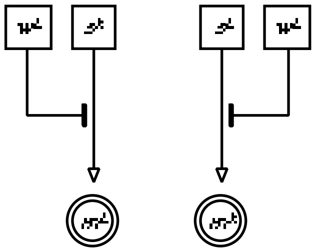


Figure 4.2.17 Neutral avoidance. The signals from the two whiskers suppress the signals from the smell sensors to make the creature stop before it collides with an obstacle.

When the potential field method is used for path planning, obstacles are considered as aversive objects with a very steep avoidance gradient, as in Hull's classical analysis. Avoidance gradients from different obstacles are added together to yield the behavior of the robot. This scheme has the disadavantage that the robot will avoid two obstacles twice as much as a single obstacle. A larger collection of obstacles may even generate active avoidance behavior, that is, it seems as if the robot is afraid of obstacles. If the neutral gradients are added together, the same problem can occur. We can, thus, conclude that neutral avoidance gradients should not be composed using addition. What mechanism should be used instead? It turns out that a selection of the maximal avoidance gradient does a better job. When many obstacles are present, only the one closest to the animal will have any effect. This will also be discussed in section 4.3.

Behavior in Relation to an Unknown Object The most interesting class of objects are those that are unknown. There is an inherent conflict in the behavior that should be directed toward such object. Since an unknown object could be dangerous, it should not be approached. However, the object may be appetitive and should, thus, be approached. An unkown object is a typical case of an approach–avoidance conflict (See section 4.3). The solution to this conflict is to approach the object but with caution.

This behavior varies according to how much is known about the object. The creatures that will be described here use the strategy of approaching the unknown object from a distance but avoiding it when they come closer. This makes it possible for them to learn about the object while still being able to avoid it if it is dangerous. When learning is included in the nervous system, the avoidance tendency decreases over time if the animal is not hurt, until it is able to approach it all the way.

Figure 4.2.18 Combined active avoidance. The creature uses an approach strategy at a distance, but avoids the object at a closer range.

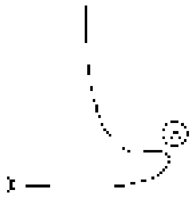


Figure 4.2.17 shows how a combined active avoidance behavior can be used to investigate an unknown object. In chapter 7, we will see how exploratory behavior is generated using a combination of behavioral inhibition, orienting behavior, and the combined active avoidance behavior. A sequence of these behaviors will be called an *orienting reaction* or *orienting response*. (See also section 4.4.)

This concludes the discussion of behavior directeded toward or away from a single object. Below, we will see examples of more complex behaviors that cannot be considered as directed in the sense that they cannot be generated by a composition of potential fields. First, however, we must investigate how an animal can behave in relation to multiple objects.

4.3 The Coordination of Behavior

In chapter 2, we saw that many behaviors are generated by innate releasing mechanisms. Such systems can be modelled by two interacting components (Figure 4.3.1). The first is an *applicability predicate* (Connell, 1990) which determines if the sign-stimulus is present and the second is a *transfer function* which transforms the sensory input to an actuator output. When the applicability predicate recognizes the sign-stimulus, it activates (or facilitates) the transfer function that consequently starts to generate an output.

The applicability predicate corresponds to a rudimentary perceptual categorization while the transfer function corresponds to a motor control system. In many animals, the influence of the perceptual system is not to *activate* a motor system but to inhibit (or supress) it. For example, an earthworm deprived of its ‘brain’ will start to crawl unconditionally (Shepherd 1988). In this case, the role of the brain is to inhibit the motor action when it is not needed. Logically, the two cases are identical.

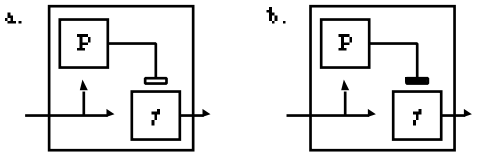


Figure 4.3.1 The internal structure of a behavior module. (a) An applicability predicate, P, determines if it is appropriate for the module to produce an output. If this is the case, the control output is generated from the input by the transfer function *f*. (b) The applicability predicate determines when it is appropriate to inhibit the transfer function.

Note the relation between a behavior module and structures of the type <precondition, action> that are used in classical symbolic AI. The main difference between a behavior module and such a structure is that a behavior module is a fixed, and usually physical, unit that includes a control strategy with the action command. See Nilson (1994) for an example of a symbolic structure that is very similar to a behavior module.

More complex action patterns can be generated if many behavior modules are connected in a sequence within an engagement system (See section 3.2). There are four distinct ways to do this depending on two variables. The first choice is whether the linking between the modules is internal or external, and the second depends on whether the behavior modules use sensory feedback or not (figure 4.3.2).

For a *fixed-action pattern*, the function *f* generates a time-dependent actuator output that is independent of any sensory input. As a result, the sign-stimulus sets off a sequence of actions that are executed without any regard for their consequences in the world (figure 4.3.2a). An action sequence of this type is thus *internally-linked* and does not use feedback. This is the type of behavior that was associated with a response chain in section 2.7.

Internally-linked sequences have the problem that they generate the whole sequence of movements, even if this does not have the expected impact on the external world. This problem can be overcome with the help of *externally-linked* behavior modules (figure 4.3.2b). In this case, the behavior modules communicated with each other *through the world* (Brooks 1991a). The first behavior module generates an action that has consequences in the world which are detected by the second behavior module in the sequence, and so on. If the outcome of a behavior is not the expected one, the action chain will be terminated, or a jump will be made to a more appropriate place in the sequence. The movements are still rigid, however, since no sensory feedback is used. In section 2.7, this type of behavior was assumed to be generated by an S–R chain. Brooks (1991a, b) has made a strong case for the use of externally-linked control systems for robots.

To generate better behavior, both internally and externally-linked behaviors can make use of sensory information, both in the form of perception and proprioception. Figures 4.3.2c and d show two architectures of this type. The stimulus–

approach sequence that was introduced in section 2.7 is an example of a behavior sequence that uses feedback in each step.

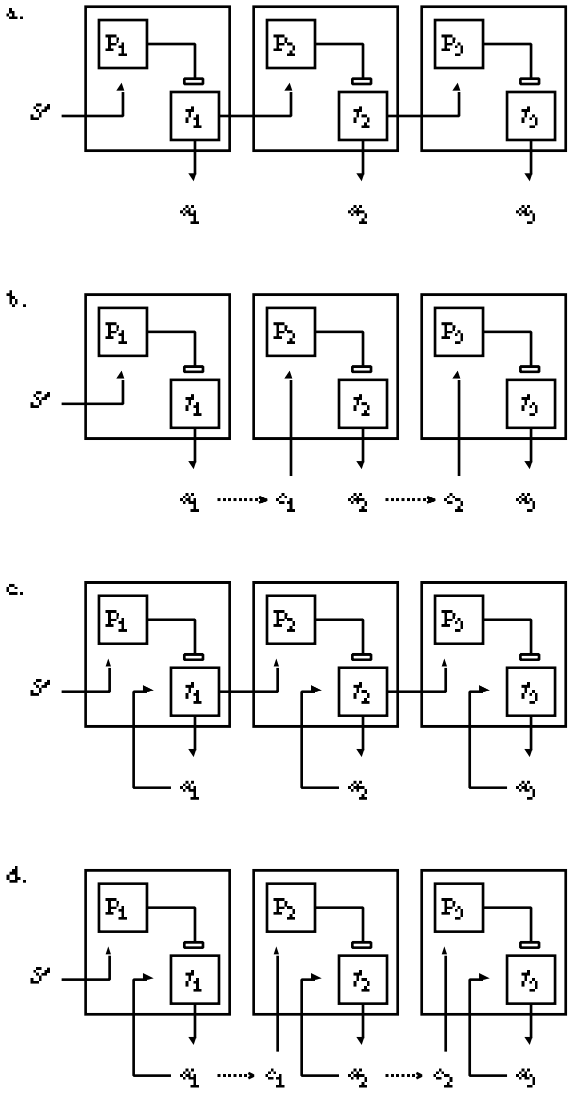


Figure 4.3.2 Four types of action sequences. (a) Fixed-action pattern: a sequence of actions a_1, a_2 and a_3 is generated without any sensory feedback when the sign-stimulus S^s is detected. (b) Hierarchical action pattern communicating through the world. Each behavior module produces an action a_i which has certain consequences c_i in the world. These in turn trigger the next behavior module. (c) Fixed behavior sequence where each behavior module uses sensory feedback. (d) Hierarchical action sequence with feedback. Species-specific drive actions are typically of this kind.

In real animals, externally-linked behaviors are best known through the work of Tinbergen (Tinbergen 1951/1989), who called them *hierarchically arranged instincts* or *hierarchical motor patterns*. A related concept is that of a *species-specific drive action* (Lorenz 1977, p. 57). An important aspect of such sequences is that each behavior leads to a new and more specific situation where a new sign-stimulus can be found. For example, a bat's search for insects can be divided into a distinct set of phases in this way (Simmons 1989). The first phase consists of *search* for a prey. When a prey is detected, a second phase starts. At this point, an *approach* behavior commences and continues until the bat is within approximately one meter of the target. Here the third phase begins, and the bat begins to *track* the insect. Finally, a *terminal* behavior is executed which leads to the capture of the prey. In the bat, the various phases of the appetitive behavior can be distinguished both by the behavior of the animal and by the sound signals it emits to locate its prey (Compare section 2.2).

Action Selection

In this section, we will consider a number of different ways in which behaviors can be composed. In behavioral robotics, this selection process is often refered to as *arbitration* or *action selection*, and the nodes in the network of behavior modules that are responsible for the combination are called *arbitration nodes*. The role of these systems is to control the coordination of different actions.

The concept of action selection will be the basis for the theory of motivation that will be developed in chapter 6. Here, we will only consider the most basic forms of action selection and behavior composition which are needed by a reactive agent. The different types of arbitration methods are summarized in figure 4.3.3 below.

Let us start with the *additive composition*. Two behavior modules generate control signals that are simply added together. As we have already seen, this is the mechanism that is mostly used to combine potential fields (Arkin 1990). A nice property of this method is that it does not matter whether we sum the potentials or the gradients to get the composed behavior. This follows from the fact that is a linear operator. Thus, if $G_i(z)$ is a set of potential reward functions,

$$\sum_i G_i(z) = \sum_i \quad G_i(z) \qquad (4.3.1)$$

The left hand expression corresponds to the situation where the individual potential reward functions are first added together before the direction of movement is calculated. The right hand expression corresponds to the case where each behavior module generates its own actuator output which is subsequently added together with the other. Since the individual behavior modules may include an applicability predicate that controls whether the module generates an output or not, no additional arbitration is necessary in many cases.

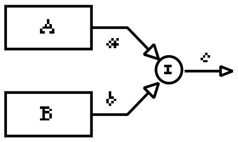
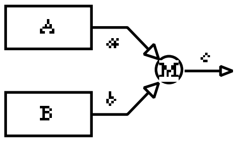
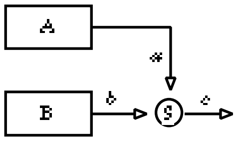
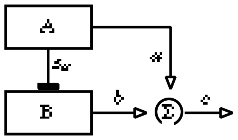
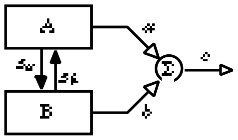
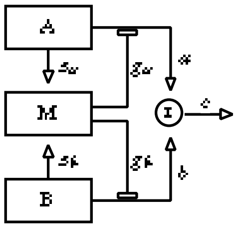
Type	Diagram	Arbitration
Additive		$c = a + b$
Maximum		$c = \begin{cases} a & if\ \lvert a \rvert > \lvert b \rvert \\ b & otherwise \end{cases}$
Suppression		$c = \begin{cases} b & if\ a = \mathbf{0} \\ a & otherwise \end{cases}$
Inhibition		$c = a + (1 - s_a)b$
Mutual Inhibition		$c = (1 - s_b)a + + (1 - s_a)b$
Central Selection		$c = g_a(M)a + g_b(M)b$ $M = m(s_a, s_b)$

Figure 4.3.3 Summary of the different arbitration methods. The vector-valued actuator commands from behavior modules A and B are called a and b, respectively. The output from the arbitration node is called c. Strength outputs are called s_a and s_b and gating signals are called g_a and g_b. In central selection, the gating signals are functions of a central motivational state M. This state, in turn, depends on the strength output from A and B.

One drawback with additive composition is that it generates unsuitable behavior in some cases. If there are a number of goals close to the creature in the environment, it will appoach the average position of the goals instead of one distinct goal. This is obviously not a very good strategy although it does manifest in real animals.

An animal can overcome the problem of multiple goals by using *maximum composition*. In this arbitration scheme, only the control signal with the largest magnitude is allowed to control the creature. This is similar to response competition which has been an important part of many theories of behavior (Hull 1934, Grelle 1981). Note that, in general, the control signals are vectors and not scalars. As a consequence, it is necessary to look at the magnitude of the whole vector and not only at the individual components when a control signal is selected.

The *subsumption* architecture invented by Brooks (1986) is yet another variation on this theme. In this architecture, arbitration depends on a fixed hierarchy of the control signals. The arbitration nodes are called *suppressor* nodes. When the behavior module at the top generates an output, it supresses the control signals from the lower behavior module. This situation is shown in figure 4.3.3. The behavior-based robots built by Brooks (1990) and coworkers) show that it is possible to construct quite complex systems using an arbitration scheme of this type.

Supression can alternatively be seen as a form of *inhibition*. Inhibitory signals can either decrease the intensity of the output of the other module or can be used to inhibit it completely. This second situation can be interpreted as if the applicability predicate of the behavior module at the bottom depends on the output of the top module.

It is also possible for the behavior modules to mutually inhibit each other. When *mutual inhibition* is used, it is possible for the behavior modules to compete in such a way that only one of them will be able to generate an output signal. In the neural network literature, this is usually refered to as a “winner take all” architecture (Compare section 2.3). In many respects, this arbitration scheme is similar to maximum composition, but with one important difference. When maximum composition is used, it is the magnitute of the output signals that determine which behavior will win. With mutual inhibition, this need not be the case. It is quite possible for the inhibitory output to be very strong and the control signal to be very weak at the same time. This makes it possible to have a very strong tendency to do very little. The signal that represents the strength of the behavior module will be called the *strength output* in order to distinguish it from the control output (See figure 4.3.3).

An even more complex behavior selection scheme can be called *central selection*. It is similar to mutual inhibition in that the different behavior modules compete for activation. Another similarity is that the signal that determines the strength of each behavior tendency is independent of the control signal. A central motivational module is used to collect the strength outputs from each behavior module and select which control signals should be let through the arbitration node. In chapter 6, we will study how the strength outputs can be used to represent the estimated potential reward that will be received if the behavior module from which they emanate are activated. We will argue that the *motivational system* is based on this type of arbitration. In this case, the strength output from the behavior modules are used as incentive signals (See chapter 6).

Finally, it is possible to select action according to a stochastic function. In *random selection*, the probability for a certain action is proportional to its strength. If $a_0 \ldots a_n$ are a set of behaviors generated by the behavior modules $A_0 \ldots A_n$, the probability that A_i will be performed is calculated from, for example, a Bolzmann-distribution as,

$$p(c = a_i) = \frac{e^{|a_i|/T}}{\sum_{j=0}^{n} e^{|a_j|/T}} . \qquad (4.3.2)$$

The parameter T is called temperature and controls the randomness of selection (Ackley, Hinton and Sejnowski 1985). The probability of selecting the largest signal increases with decreased temperature. When T approaches infinity, the arbitration becomes entirely random. When it is close to 0, the behavior with the largest strength is always selected. As in the case of central selection, it is possible to let strength outputs control the choice instead of the actual behavior signals. Random selection is, thus, a form of central selection since the choice of behavior must be based on strength inputs from all behavior modules.

Behavior in Relation to Multiple Objects

In any realistic situation, the environment consists of a large number of objects, some of which are appetitive and some which are aversive. Those objects which are appetitive and those which are aversive may change over time as the needs of the animal change. At a single moment, however, the potential rewards corresponding to each object in the environment can be considered constant. In Lewin's theory, this is called the *valence* of an object (Lewin 1935). The valence is responsible for the strength and direction of the forces around an object. An object with a positive valence will attract the individual, while an object with negative valence will push the individual away from it. The forces from all the objects in the environment are combined to describe the behavior of the animal. This clearly corresponds to the addition of potential fields which was discussed above. In this section, we will consider the dynamics of situations with more than one object. The presentation will start with some classical examples of conflicts.

The Approach–Approach Conflict Consider a runway with food placed at each end (figure 4.3.4). A hungry rat that is placed in the middle of this runway has the same distance to both pieces of food and is, thus, in a state of conflict. The potential reward function for the situation does not tell the animal in which direction to move. This is called an approach–approach conflict (Lewin 1935). Two approach tendencies compete with each other in such a way that they neutralize each other. The resulting behavior of the animal, if it were to follow the potential reward function, would be to remain at its current location. However, any movement toward one of the goals would make the force toward that goal stronger and toward the opposite goal weaker, and the animal would start to approach it.

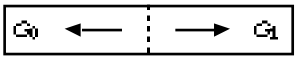


Figure 4.3.4 The approach–approach conflict in a runway with a goal object at each end.

As pointed out by Miller (1944), the point in the runway exactly in the middle between the two goals corresponds to an unstable equilibrium point. This equilibrium holds for all points along the line that divides the runway in the middle. This line is called the *Voronoi border* between the two basins of attraction for the two goal objects. Figure 4.3.5 shows the Voronoi borders of an environment which consists of 3 objects with equal positive valences. In general, the different object may have different valences, and the borders between the basins of attraction will not be at equal physical distance from the goal objects.

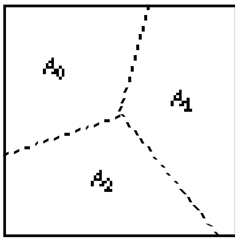


Figure 4.3.5 The general approach–approach conflict occurs at the Voronoi borders between the basins of attraction around different goal objects.

Figure 4.3.5 illustrated the basins of attraction as they are generated by the potential reward functions of the environment. When the animal tries to follow this function based on its momentary sensory input, the situation looks a little different. Figure 4.3.6, shows how a creature that uses a combined approach behavior will behave in an environment with four goal objects. The simulated creature is placed at each location in space and is allowed to move until it reaches a goal. If it reaches the goal at the top left, the starting position is marked by a black dot; if it reaches any other goal first, the starting position is marked in white. The creature is initially oriented toward the center of the environment. This means that the figure only shows one projection of all the possible initial locations and postures.

As can be seen, the creature fails to follow the potential reward function for many initial locations. It appears that the initial orientation of the creature plays a larger role in determining its goal than the distance to the closest goal. As can be expected, a creature that first orients toward the closest goal will behave more in agreement with the potential reward function.

Figure 4.3.6 (a) The environment. (b) The basins of attraction in a simulation of combined approach. (c) Unsafe points.

The simulation shows that there exist locations very close to each other which result in different goals. For a given uncertainty *s*, let us define a location as *unsafe* when there is another location at a distance less than *s* that results in a different goal object (See Peitgen, Jürgens and Saupe 1992). Figure 4.3.6c shows the set of unsafe locations for the situation in figure 4.3.6a. The area of these points can be considered as a measure of the width of the boundaries between the basins of attraction for the two goals.

In Hull's analysis of the approach–approach conflict, he thought it necessary to include an oscillation function that would generate the required deviations from the unstable equilibrium seen in animal experiments. In the simulation shown here, deviations from the equilibrium points are the result of deterministic chaos. This has some important consequences.

First, since the choice of goal is sensitive to the initial location of the creature, it is not, in general, possible to predict which goal the creature will approach with any certainty. Note that this is not a result of any randomness in the behavior but is an inherent property of many environments with multiple competing goals.

Second, the boundary between different basins of attractions can not be considered as a line, but must be seen as a region. Usually, such a region is of fractal dimension. A further discussion of these properties would, however, lead us too far away from the topic of this book. The interested reader is instead referred to the clear-cut presentation of the relation between dynamical systems, chaos and fractals that can be found in Peitgen, Jürgens and Saupe (1992).

The Avoidance–Avoidance Conflict Another type of conflict situation occurs when the objects at each end of the runway are aversive. Given that the creature uses active avoidance behavior away from each of the two aversive objects, it will eventually end up in the middle of the runway. In the ideal situation, this location constitutes a stable equilibrium (figure 4.3.7). In a real situation however, the heading of the creature will influence the two potentials and it will oscillate around the middle of the runway.

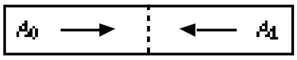


Figure 4.3.7 The Avoidance–Avoidance Conflict

The Approach–Avoidance Conflict The final conflict situation is the approach–avoidance conflict that results when a location is both appetitive and aversive at the same time (figure 4.3.8). We have already seen that this conflict is present when an animal encounters an unknown object. However, situations of this type can also be artificially produced, for example by placing both an appetitive and an aversive object at the same place (Lewin 1935, Miller 1944; see also figure 4.2.17).

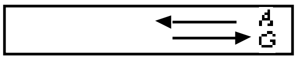


Figure 4.3.8 The Approach–Avoidance Conflict

4.4 An Elementary Reactive Repertoire

In this section, we will develop a basic set of reactive behaviors, most of which will not be goal-directed. To behave appropriately, a creature will need a large set of such behaviors. The repertoir we will consider here is only meant as a rudimentary example. It shows the diversity of the innate behavior modules that must be present before an animal can do anything at all. Most of all, the innate behavioral repertoir makes it possible for the animal to stay alive long enough to make use of its learning abilities. These behaviors can be seen as reactions to problems that are certain to be present in the environment, that is, problems that do not generally gain any advantage from learning.

Grooming

Grooming behavior will be generated by a set of nodes that activate each other in sequence. Each node moves the whiskers to a certain position, and the sequence of activation will, thus, generate a stereotypic motion of the whiskers. A network architecture of this type has been called an avalanche since it cannot be stopped once it has started (Grossberg 1986). The network in figure 4.4.1 shows an avalanche of this type. When the integrating node to the left receives an input signal, it start to oscillate as described in section 2.3. Every time its output signal is high, it will start the avalanche which in turn produces a stereotypical action pattern. This is, thus, a simple case of an internally-linked behavior.

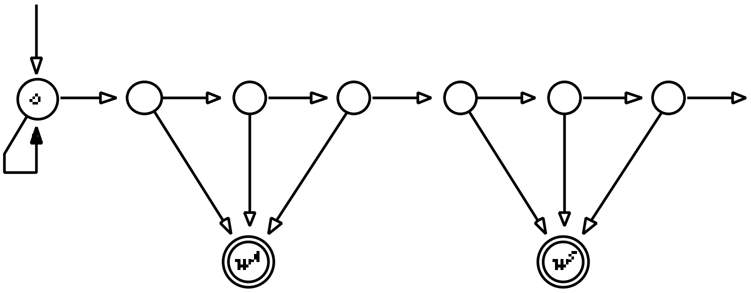


Figure 4.4.1 An avalanche triggered by an oscillator produces a stereotypical action pattern.

Such behaviors are very common in real animals. “Typically, rats lick their fur, then scratch themselves, and then fiddle with their feet. The different components usually appear in this order.” (Bolles 1984, p. 436) Grooming is obviously much more complex than the behavior generated by the network here, but the notion of an avalanche does, nevertheless, capture an essential feature of such behavior.

The Orienting Reaction

The orienting behavior of our creatures will consist of three phases. The first is the inhibition of any ongoing behavior by activation of the behavioral inhibition system (See figure 4.2.10). This is usually called *external inhibition* (Pavlov 1927) in order to distinguish it from internal inhibition which will be discussed in chapter 5. In the second step, an orientation movement will be performed toward the stimulus using the method described in figure 4.2.11. Finally, an exploratory behavior will be executed. Such a behavior will typically be a combination of an approach and an avoidance behavior. The stimulus will be approached from a distance, but avoided when the creature comes closer to it, as shown in figure 4.2.17. The activation of these three phases can be controlled by an avalanche in a manner similar to the grooming behavior above.

There is one problem with this orienting behavior which we have already mentioned. When many stimuli are present in the environment, the creature may sometimes orient toward the average position of them and not to one of them. One solution to this problem is to let the creature use a spatiotopic representation of the environment. In such a representation, a specific node is allocated to the representation of every location in the environment around the creature. Since each stimulus is mapped onto its own node, or set of nodes, the locations will not blend into each other. This is the type of representation that is used within the orienting system of many real animals (Stein and Meredith 1993). Such a system resides in the superior colliculus in mammals and the homologous optic tectum of birds and reptiles.

This more complex type of orienting system will not be used here for two reasons. The first is that it requires a very large network which would make our creatures unnecessarily complex. The second is that a spatiotopic representation is more suitable for a creature that uses other modalities such as vision and hearing. A neural network model of this type of orienting system is presented in (Balkenius 1995; see also chapter 9).

Reactive Defense Mechanisms

Aversively motivated behavior will play a rather small role in this book, but it will, nevertheless, be useful to consider some basic behaviors of this kind. It has already been suggested that it is possible to model some defense mechanisms by a behavioral gradient that makes the creature freeze when the aversive stimulus, such as a predator, is far away and flee when it is closer. While this type of system is a very crude analogy to the defensive behaviors of real animals, it shows how the intensity of the stimulus gradient around an object can be used to select different behaviors. This is, thus, a kind of arbitration mechanism.

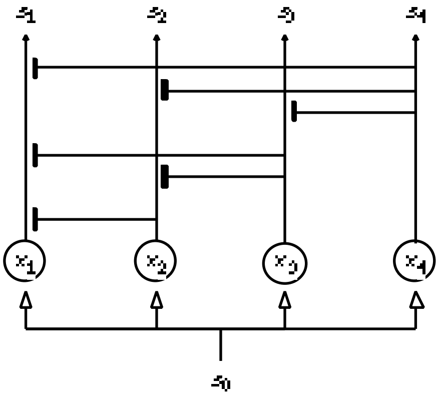


Figure 4.4.2 A network that selects an output $s_1 \ldots s_4$, depending on the level of the input signal s_0.

Figure 4.4.2 shows a network that is able to divide a signal into intervals in this way. Each node has a threshold output function, and the threshold increases with increased index on the node. Since each node supresses the output of the nodes with smaller index, the level of the input signal s_0 is used to select the output. This network could be used to select the different behaviors depending on the level of fear. A network of this type could form the basis for a fight, flight, or freeze system (Cannon 1932). A similar network could be used to select between offense, defense, or submission depending on the estimated strength of an opponent (Adams 1979). A related function is the choice between an orienting reaction or a defensive reaction depending on the intensity of a stimulus (Raskin 1972, Stein and Meredith 1993). For example, a sound with low amplitude will activate an orienting reaction toward it while a sound with high amplitude will generate a defensive reaction away from it.

Wandering

To navigate succesfully in a more complicated environment, potential fields are not sufficient on their own. There are many behaviors that are useful but cannot be generated from a potential field. In this section, we present a number of behaviors that are useful in the spatial domain, but must be generated in a way that differs from those we have seen so far. These behaviors are not goal-directed and we will refer to them as *wandering*. Wandering behaviors include, for example, wall-following, corridor-following, obstacle avoidance, and door entrance. Usually, behaviors of this type increase the chances of achieving consummatory behavior, and they can, thus, be considered as appetence behaviors. Wandering can also be considered as undirected exploration (Thrun 1992).

Random Walk In the final example of the previous chapter, we have already seen how to construct a behavior that will move the creature around more or less at random on an empty surface. Such a behavior is often called a *random walk*. Since such a random walk is not directed at all, it is obvious that it cannot be generated by a potential field and this is, thus, our first example of a behavior of this kind. When the environment is structured, it is often advantagous to use more complex behaviors.

Wall-Following When an animal senses a wall, it is very often useful to be able to follow this wall instead of turning away from it. This is a common behavior of many insects, and for the maze-like environment of our artificial creature, it will obviously be of great utility. We may connect wall-following to the concept of affordance and say that walls afford following (Gibson 1979). Whether the creature will follow the wall or not may depend on its current needs and if wall-following has been rewarded or not.

To accomplish wall-following, all our creature has to do is move in a direction parallel to the wall. To do this, it must use its whiskers to feel the distance to the wall and try to keep this distance constant.

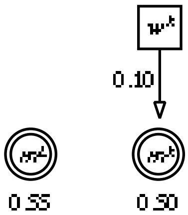


Figure 4.4.3 A network that produces wall following of a wall which is to the right of the creature. When the creature receives no external stimulation, it will turn toward the right since the left motor neuron has a larger resting activity than the left. When the right whisker comes into contact with the right wall, the right motor will receive a larger input signal that will cause the creature to move away from the wall. As a result, the creature will follow the wall.

Figure 4.4.3 shows a nervous system that will sucessfully follow a wall to the right of the creature. When the sensory signals are above a set-point, the creature will turn away from the wall. When the whisker signal is below this value, the creature will, instead, turn toward the wall. As a result of these two opposing forces, the creature will move approximately parallel to the wall (figure 4.4.4a). This is, thus, an example of a servo–mechanism (Gallistel 1980). An identical network can be constructed which lets the creature follow a wall on its left.

An important aspect of the wall-following behavior generated by the nervous systen in figure 4.4.3 is that it is relatively stable. Even if noise is added to the sensory signals, the creature will continue to follow the wall, albeit in a somewhat more random fashion. This is shown in figure 4.4.4b, where 20% noise has been added to the sensors. Figure 4.4.4b shows what happens when the whiskers point straight to the side of the body of the creature. In this case, the wall-following behavior is unstable even though no noise has been added.

Corridor-Following In the artificial world of our creature, and in the classical mazes, corridors are very common. The ability to follow a corridor smoothly does, thus, seems useful. Such a behavior can be generated in a way similar to wall-following, but both whiskers must be used. When the creature is in the middle of a narrow corridor, the signals from each of the whiskers are equally strong. As a consequence, the creature can follow the corridor if it tries to keep the signals from the two whiskers equal. Figure 4.4.5 shows a nervous system that can generate this behavior. A simulation is shown in figure 4.4.6. As can be seen, the creature quickly aligns itself with the two walls and starts to follow the corridor. Again, it is possible to show that the behavior is stable if the whiskers point forwards but unstable if the whiskers point straight out from the sides.

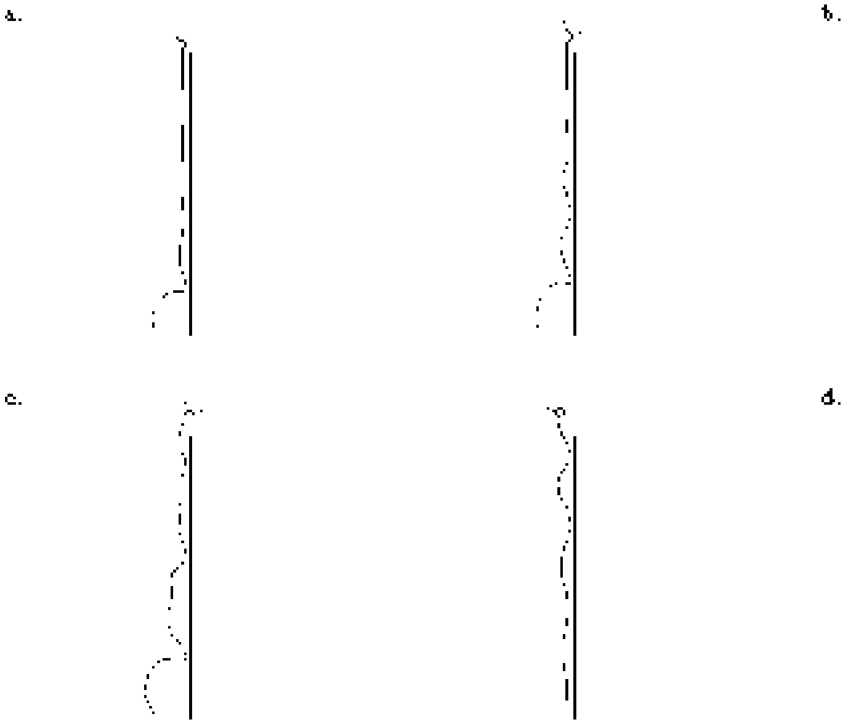


Figure 4.4.4 Wall following. (a) The behavior of the network in figure 4.4.3. (b) Stable behavior with 20% noise. (c) Stable behavior with 60% noise. (d) Unstable wall-following results even without noise if the whiskers point directly to the sides.

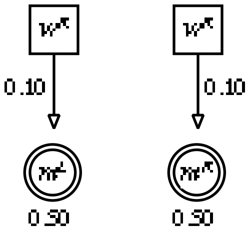


Figure 4.4.5 Network that produces corridor-following behavior in a narrow corridor. Without sensory stimulation, the creature moves straight ahead. When it comes into contact with a wall, it will make a compensatory movement away from it. If the creature can sense both walls at the same time, it will align itself with the corridor.

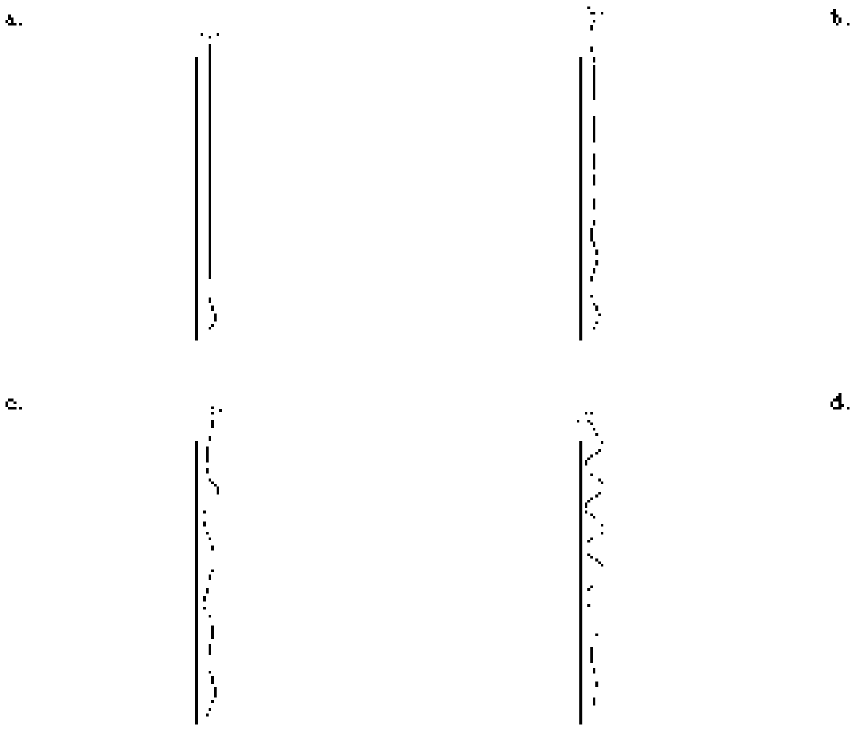


Figure 4.4.6 Corridor-Following. (a) The behavior produced by the network in figure 4.4.5. (b) Stable behavior with 20% noise added. (c) Stable behavior with 40% noise. (d) Unstable behavior results if the whiskers point directly to the sides. Although the creature begins aligned with the corridor, its behavior starts to oscillate after a while.

Obstacle Avoidance Another useful ability is to move around obstacles. In the simplest case, this means that the creature should simply turn when it senses something with one of its whiskers. A more sophisticated strategy would enable the creature to continue in its original direction when the obstacle has been negotiated. To do this, the creature must keep track of its deviation from the original path.

Figure 4.4.6 shows a simple network that accumulates deviations from the original direction of movement. When the obstacle avoidance starts, the whiskers control the movement away from the obstacle. During this operation, the node, R, is supressed by the whiskers and the the nervous system accumulates the deviation from the original direction of movement in an integrating node called D.

When the obstacle has been avoided, the node R is disinhibited, which lets the accumulated deviation in D direct the creature back toward its original path through the connection α. The connection weight α determines how fast the creature will return to its original path. A computer simulation of the behavior produced by this circuit can be found in figure 4.4.7.

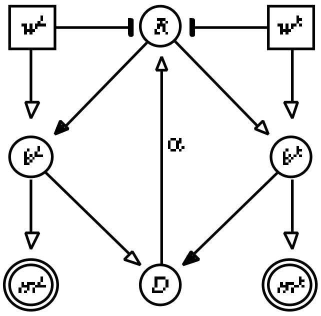


Figure 4.4.7 A simple network for obstacle avoidance. When the creature turns away from an obstacle, the node *D* will accumulate the deviation from the original direction. When the obstacle has been negotiated, node *D* will turn the creature back on its original path through node *R*.

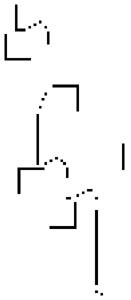


Figure 4.4.8 Computer simulation of the network in figure 4.4.6. When the creature senses an obstacle, it turns away until its whiskers stops reacting. When this happens, the creature tries to return to its original heading.

The signals of the type accumulated in *D* are somtimes called efference copies (Gallistel 1980). They are copies of the signals sent to the motor system. The accumulation of turning signals is, thus, a form of dead-reckoning (Gallistel 1990). A similar mechanism can be used for path-integration to update the location of the creature as it moves around in the environment (Gallistel 1990, Touretsky and Redish 1995).

When the creature is performing an approach behavior, obstacle avoidance is much simpler. The perception of the goal will automatically turn the creature in its original direction again when the obstacle has been avoided.

Choice Each of the networks presented above in this section should be considered as a behavior module. It is clear that many behavior modules of these types must be combined before some interesting wandering behavior can occur. The first requirement is to add an applicability predicate described in section 4.3 which determines when it is appropriate to perform each behavior. For example, the wall-following behavior should only be induced when a wall is present. More generally, it is possible to use the various arbitration mechanisms described in section 4.3 to choose between the different behaviors.

4.5 Conclusion

We have seen how an animal can react to four classes of stimuli: appetitive, aversive, neutral and unknown. Each of these stimuli are encountered during either appetitive, aversive or orienting behavior. In an appetitive situation, the behavior of the animal makes it likely that it will find some desired object such as food. Figure 4.5.1 summarizes the main components of the appetitive behavior of our creature. Note that neutral avoidance is only encountered during appetitive behavior as a special case. It is goal-directed in the sense that the location of the goal controls the behavior, but it is also similar to wandering since the avoidance is not goal-directed.

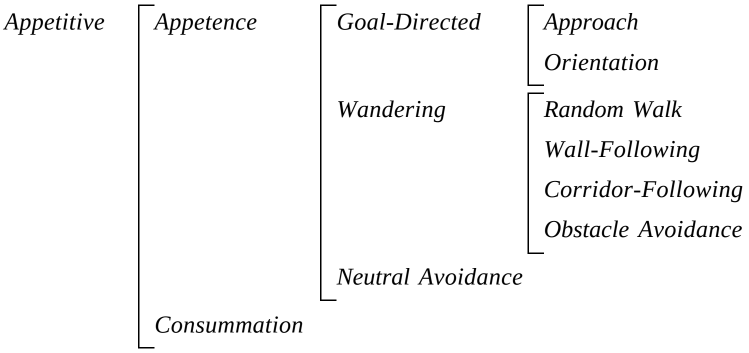


Figure 4.5.1 Summary of the behaviors used in an appetitive situation.

The different types of aversively–motivated behaviors are shown in figure 4.5.2. Note that passive avoidance can only be used in combination with some appetitive behavior since its only function is to inhibit such a behavior.

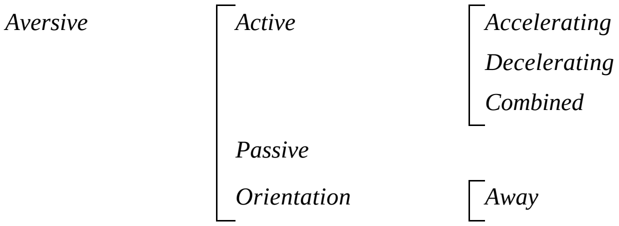


Figure 4.5.2 Summary of aversively motivated behaviors.

In the final class situation that we have considered, the animal is required to investigate a novel stimulus. The behavior produced is called an orientation reaction and consists of the three components shown in figure 4.5.3.

Orienting Reaction
- *External Inhibition*
- *Orienting Toward*
- *Approach with Caution*

Figure 4.5.3 The components of the orientation reaction

The orientation reaction is different from appetitive and aversive behavior in that it must exist together with some learning ability. Without learning, the creature will never retain anything about a stimulus, and, thus, every experience will be novel. As a consequence, the creature will never stop its orientation reaction. In the next chapter, we will see how learning can influence this reaction and extinguish it when the stimulus is no longer novel. We will also see how learning can be used in both the appetitive and the aversive situation.

Chapter 5

Modeling Learning

5.1 Introduction

In this chapter, we will take a look at the most fundamental types of learning and analyse how they interact with each other to produce adaptive behavior. The presentation will start with the two major types of learning experiments that traditionally have received the most attention: classical and instrumental conditioning. We will then investigate, how these two experiment types relate first to learning mechanisms, and then to each other. We will see that both types of learning can be based on the same mechanism. This learning mechanism, which we will call a *reinforcement module*, is an important component in the different engagement modules using learning. The model we will develop is similar to earlier two-process models of conditioning (Mowrer 1960, Gray 1975, Klopf 1988) and also share many properties with other reinforcement learning schemes such as Q-learning (Watkins 1992) and temporal–difference learning (Sutton and Barto 1990).

The next step is to consider some basic engagement modules and how they exploit learning. We will investigate learning in the appetitive food-related situation, learning of escape, learning of action sequences, and the learning of expectations. Perceptual learning will be discussed in chapter 7; and in chapter 8, the mechanisms described below will be used for learning in the spatial domain.

Our argument is, thus, that many types of learning can be based on similar mechanisms, and not that one learning mechanism solves all problems. Since learning is only useful when it is embedded in a specific engagement module, the reinforcement module in itself should be considered as an abstraction.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

5.2 Instrumental Conditioning

The study of instrumental learning is usually associated with Thorndike. In one of his well–known experiments, he placed a cat into a box where it was required to press a bar or drag a string in order to open the door and get out. When the cat eventually succeeded in escaping, it was rewarded with food or water. Thorndike noted that for each time the cat was put in the box, the faster it managed to escape. To explain why the cat performed better for each trial, Thorndike suggested that the reward given after the cat had escaped the box would gradually strengthen those behaviors performed immediately before the reward was presented. As a consequence, behaviors that let the cat escape would increase in probability while behaviors that had no effect would become less likely the more practice the cat received. This property was formulated as the 'Law of Effect', which states that the effect or a behavior determines whether this behavior will be performed again or not in the same situation.

The important property of instrumental learning is that learning depends on the *consequences* of a behavior performed by the animal. A learning trial can be described by three components: (1) the situation perceived by the animal, (2) the behavior performed by it, and finally, (3) the consequence of performing the behavior. The consequence is said to reinforce an assumed connection, or association, between the situation and the behavior. When the consequence is positive, the connection is strengthened, that is, positively reinforced. When the consequence is negative, the connection is weakened, that is, negatively reinforced.[1]

A classification of different types of behavioral learning experiments was presented in Gray (1975) and are reproduced in figure 5.2.1. This classification has its origin within the behavioristic tradition and should, thus, be seen as descriptions of experimental procedures rather than learning mechanisms. Since mechanisms are the primary interest here, this piece of historical luggage will not bother us here. We will see that there are reasons to believe that the different types of experimental procedures do, in fact, make use of different, but interacting learning mechanisms within an animal.

The different instrumental learning types can be distinguished by two factors. The first is the procedure used, and the second is the outcome it produces. The result of the event can be either that the probability of a behavior preceding the presentation increases or decreases. When the probability of a behavior increases, the

[1] In some theories of conditioning, the terms "positive" and "negative reinforcement" are used in a different way than here. As we use the term "negative reinforcement", it refers to the strengthening of an *inhibitory* connection. In Skinner's theory it refers to the strengthening of an *excitatory* relation between stimulus and response caused by the termination of an aversive state. In the model presented here, the term *reinforcement* can be equated with learning signals within the artificial nervous systems, whether or not its effect is excitatory or inhibitory.

event is said to generate positive reinforcement. When the probability of the behavior decreases, the reinforcement is called negative.

The simplest event is the presentation of a stimulus. Depending on the consequences of the presentation, the stimulus is either called rewarding (*Rew*) or punishing (*Pun*). If the stimulus is rewarding, the animal is more likely to do what ever it did before the presentation of the stimulus. If it is punishing, the animal will instead be less likely to reproduce the behavior that preceded the presentation. When the reward or punishment is directly derived from an external stimulus without prior learning, it is called *primary*. When its reinforcing properties depend on a previous learning experience, it is called *secondary*.

The second type of event consists of the termination of a stimulus. The termination of a stimulus that is rewarding (*Rew*!) will act as a negative reinforcer and decrease the probability of the behavior it follows. The termination of a punishing stimulus (*Pun*!), will consequently act as a positive reinforcer.

The final type of event, the omission of an expected stimulus, is more complex. On the surface, this situation appears to be very similar to the termination of a stimulus. The omission of a rewarding stimulus ($\overline{Rew}$) is negatively reinforcing and the omission of a punishing stimulus ($\overline{Pun}$) is positively reinforcing. This type of event, however, differs from the two types presented above. Omission is not generated external to the animal. It is quite possible that a stimulus is omitted when nothing at all happens around the animal; and it, thus, seems that this type of learning requires a cognitive explanation, that is, the learning of expectations.

	OUTCOME	
PROCEDURE	$p(R)\uparrow$	$p(R)\downarrow$
Presentation	Rew	Pun
Termination	*Pun*!	*Rew*!
Omission	$\overline{Pun}$	$\overline{Rew}$

Figure 5.2.1 Summary of the different classes of reinforcing events in terms of procedure and changes in probability of a behavior (or response), *R*. (Adapted from Gray 1975).

Learning of the types described in figure 5.2.1 are alternatively called *reinforcement learning* (Sutton and Barto 1990, Watkins 1992), *instrumental conditioning* or *instrumental learning* (Mackintosh 1983). In some cases, it can also be called *operant conditioning* or *operant learning* (Skinner 1974). They all have in common that a reinforcing event which comes after behavior will change the probability of that behavior.

Figure 5.2.2 shows a neural network that can calculate the appropriate reinforcement signals from primary reward and punishment. The task for the network is to establish the appropriate connection between a contextual representation (*CX*) and an activating signal (+) and an inhibiting signal (–). In general, the contextual representation includes both the environmental situation and the current motivational state (See chapter 6).

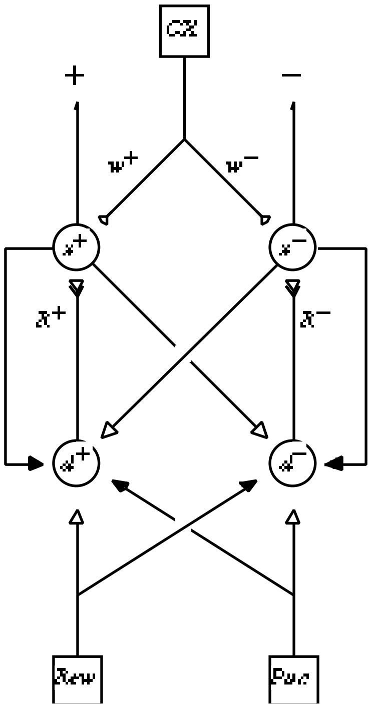


Figure 5.2.2 A network that calculates reinforcement from primary reward and punishment. All weights in this network are assumed to be 1, except for the plastic weights at w^+ and w^-.

The output of the network is assumed to control a behavior module of some of the types described in the previous chapter. If the network receives a reward when the controlled behavior is performed in the context, *CX*, the network learns to activate the behavior module at later occasions when the same context is present. If the network receives punishment, the behavior is inhibited instead. The history of reward and punishment is recorded at the connections between the contextual input and the activating (x^+) and inhibiting (x^-) nodes. The weights on these connections are called w^+ and w^-, respectively, and are changed by the two reinforcement signals that are called R^+ and R^- (A detailed description can be found in appendix A).

Let us now investigate how the conditions in figure 5.2.1 will influence the connection weights in the reinforcement network. First assume that the contextual cue (*CX*) is present and that the network receives a reward (*Rew*) for the first time. The reinforcement node (d^+) will get activated and will generate a reinforcement signal (R^+) that causes the weight w^+ to grow larger. When this happens, the activity of node x^+ will gradually increase and will, consequently, inhibit node d^+. This will make the positive reinforcement decrease until the weight w^+ assumes the same value as *Rew*. We see that the node d^+ calculates the difference between the actual reward Rew and the expected reward coded in w^+. When the actual reward is larger than the expected one, a positive reinforcement signal will be generated which changes the value of w^+. Since the reinforcement network is symmetrical, the steps involved in punishment will be the same as for reward, but in the other half of the network and the value of w^- will increase instead.

A more interesting case occurs when the contextual cue (*CX*) is present, but an expected reward is omitted. Since a reward is expected, the value of w^+ is greater than zero. This will activate x^+, which in turn sends a signal to d^-. Since no reward is received, d^- will not be inhibited, and the event will, thus, act as punishment and generate a negative reinforcement signal. This signal increases the value of w^- until it cancels out the effect of w^+. Again, the mechanism for omitted punishment is analogous and, thus, causes positive reinforcement to be generated when an expected punishment is omitted.

Termination can be seen as a special case of omission since the presentation of reward or punishment will cause the corresponding weight to increase, thereby causing the current signal to become expected. When it is terminated, the situation is exactly the same as for omission.

Note that it is the difference between the actual and expected reward or punishment which drives the learning process and not the reward or punishment in itself. As we will see below, this property has some important consequences.

Another noticable aspect is that learning only proceeds in one direction. The weights w^+ and w^- can only increase. Given that all behaviors have some effect on the environment, this has the important consequence that the network will have different representations for a behavior that it has never tried and one that it has tried when the net reward has been zero. In both cases, $w^+ – w^- = 0$, but when the behavior has been tested, w^+ and w^- are both greater than zero. It is possible to consider $w^+ + w^-$ as a measure of the confidence in the estimation of net reward $w^+ – w^-$ which will be received if the behavior is perfomed in a certain context. The network is, thus, able to distinguish between appetitive, aversive, unknown and neutral stimuli, since it has different representations for these cases (Compare section 4.2).

If $w^+ – w^- > 0$, the situation is appetitive. If $w^+ – w^- < 0$, the situation is aversive. The quantity $w^+ + w^-$ describes the novelty of the situation. If $w^+ + w^-$ is smaller than a certain threshold, the situation is assumed to be unkown (See section 7.4).

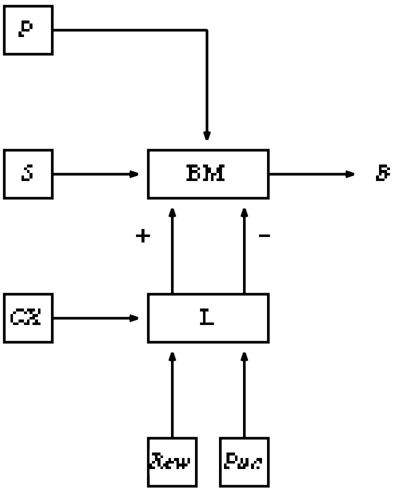


Figure 5.2.3 The components of a general reinforcement learning system. A behavior module (BM) is controlled by a stimulus (*S*) and generates a behavior (*B*) when the applicability predicate (*P*) is fulfilled. The module can be activated or inhibited by a learning system (L). The learning system, in turn, is activated when a certain contextual representation (*CX*) is present and is modified by primary reward (*Rew*) and punishment (*Pun*). Note that it is possible for *P*, *S*, *CX* and *Rew* or *Pun* to be the same stimulus.

Figure 5.2.3 presents an overview of the components involved in instrumental learning. Five types of inputs control the learning, activation and execution of a behavior. *Rew* and *Pun* control whether the behavior generated by a behavior module (BM) or an engagement module should be activated or inhibited in the future in the situation represented by the context (*CX*). The first situation is called *behavioral activation* (Gray 1995), while the second is called *behavioral* (Gray 1982) or *external* (Pavlov 1927) *inhibition*. The stimulus (*S*) is used, for example, to control the generation of a goal-directed behavior (*B*) in the behavior module. As described in the previous chapter, a behavior module is also controlled by an applicability predicate (*P*). This input makes sure that it is possible to perform a certain behavior. For instance, a behavior module responsible for wall-following should only be activated when a wall is present. This implies that a creature will not even try to learn to follow a wall in the absence of a wall.

In many cases, the inputs *P*, *S*, *CX*, and *Rew* may be the same stimulus. For example, this is often the case with a food object. First, the piece of food is used as the stimulus (*S*) which guides the approach behavior of the animal. Second, since it is appropriate to approach the food when it is present, it can also act as the context for the behavior and trigger the applicability predicate. If the food is palatable, it may finally generate a reward (*Rew*) that will make the animal approach it again at later times. In other cases, *S*, *CX* and *Rew* or *Pun* may be entirely different.

It is also possible that no stimulus is used to control the behavior generated by the behavior module. This is the case, for example, in pure stimulus–response learning, where the stimulus is used only to activate a response and not to control its execution.

5.3 Classical Conditioning

The second type of learning that we will consider is called *classical* or *Pavlovian* conditioning. This type of learning goes back to the research of Pavlov at the turn of the century (Pavlov 1927). In his most famous experiment, a dog was taught to salivate at the sound of a bell. To make the dog acquire this behavior, it was presented with food on a number of occasions after first having heard the tone of a bell. Since the bell signals that food will soon be presented, the dog learned to respond in an appropriate manner when it prepared itself for the food by salivating.

There are four components involved in classical conditioning: the unconditioned stimulus (US), for example, the food; the conditioned stimulus (CS), such as the bell that the animal learns about; the unconditioned response (UR) that is performed when US is presented; and, finally the conditioned response (CR) that is executed when the CS is presented. In the example with Pavlov’s dog, both the UR and the CR consist of salivation. Although this is still a controversial issue (McFarland 1993), it seems reasonable that the conditioned and the unconditioned response could be different.

The important difference between instrumental and classical conditioning is that in a classical learning situation, the presentation of the stimulus does not depend on what the animal does. After the presentation of the CS, the US will always follow. In the example, the dog will be given food after the bell rings whether it salivates or not.

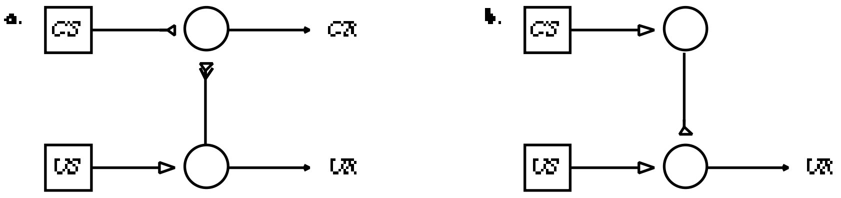


Figure 5.3.1 The two main theories of classical conditioning as simple neural networks. (LEFT) The S–R theory. When the US is presented, it reinforces the connection between CS and CR. (RIGHT) The S–S theory. A connection is formed between the representation of the CS and the US and the CS activates the response indirectly.

Two types of theories exist which try to explain the mechanisms behind classical conditioning. The first is called the S–R theory, since it suggests that what the animal learns is an association between the CS and the CR and this learning is reinforced by the US (Hull 1934). The second variation, which is called the S–S, or *stimulus substitution*, theory, proposes that what the animal learns is not an association between a stimulus and a response, but a relation between two stimuli (Pavlov 1927, Bolles 1978, Mackintosh 1983). On several occasions, the CS has preceded the US, and this is what the animal has learned. Depending on how cognitive the theory tries to be, this knowledge is either assumed to be represented by a simple connection between the CS and the US or by some more complex cognitive structure (See, for example, Pavlov 1927 and Tolman 1932).

Many versions exist of both the S–R and the S–S theory, and it is not possible to discuss all here. We will merely note that both theories have different strengths and weaknesses and it is possible that both are correct. One other thing to note is that the S–S theory in its most elementary form has problems in cases where the UR and the CR are different. A simple example where this is the case is when the dog is allowed to eat the presented food. In this case, the UR is eating, but the CR is still salivation. Very often, the CR appears to be preparatory in, which, of course, is entirely sensible. The extent to which the CR can be preparatory or not has been, and still is, in much dispute however.

If we assume that the UR and the CR (for example, eating and salivation) are produced by the same engagement module, it is possible to explain the difference in UR and CR as the consequence of different stimuli being present in the two cases. When food can be eaten, the UR (eating) will be produced; and when it can not, the CR (salivation) will be performed instead. The choice between these two behaviors can be made by the use of applicability predicates as described above. Such a mechanism enables us to adhere to a S–S theory of classical conditioning while still allowing the CS and US to be different.

An interesting approach to understanding classical conditioning was pioneered by Rescorla and Wagner (1972). They suggested that the association formed during classical conditioning should be considered as a record of how well the CS predicts the US. The main idea behind their model was to consider the association between CS and US, or possibly between CS and CR, as an *expectation* or a *prediction* that US will follow CS.

Figure 5.3.2 shows a variation of the S–R network in figure 5.3.1 which is based on the idea that the CS predicts the US. The network is similar to the instrumental learning network shown in figure 5.2.2, except that the context is replaced by the conditioned stimulus (CS) and the reward is replaced by the unconditioned stimulus (US). Like the S–R network in figure 5.3.1, it is assumed that the unconditioned response is independent of the conditioned response.

We may interpret the weight w^+ as the strength of the expectation that US will follow CS and w^- as the expectation that it will not. When US follows CS, w^+ increases, and when CS is presented alone, w^- may increase instead. To determine whether the CR should be activated or not, these two values are simply combined. If $w^+ > w^-$, the CR will be activated, otherwise it will not. Since there is no signal equivalent to punishment as in the instrumental case, it is not possible for the inhibitory signal to grow larger than the activating one, and the net activation of CR will always be positive or zero.

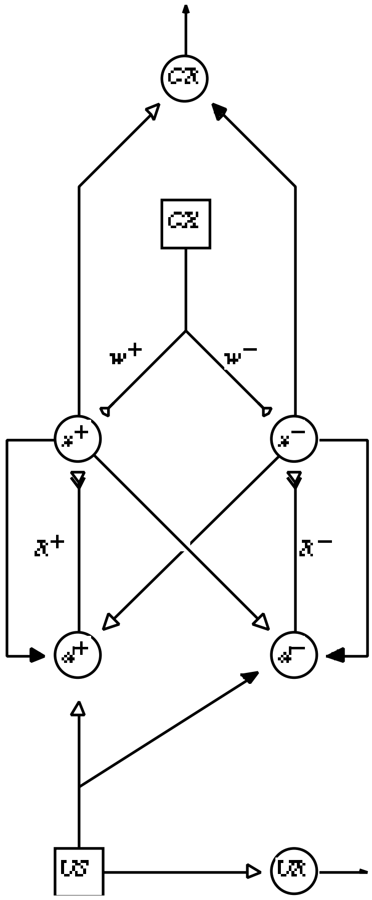


Figure 5.3.2 An expectation network for the S–R theory of classical conditioning. Note the similarity with the network for instrumental conditioning in figure 5.2.2.

In section 5.8 we will see how the same basic circuit can be used within a S–S theory of classical conditioning, but first we need to consider some more complex learning situations. Here, we will only conclude that it is possible to use similar architectures for both instrumental and classical conditioning. This insight will allow us to talk about US and *Rew* as if they were the same signal. We must keep in mind, however, that one cannot conclude that an identical physical network is involved in both classical and instrumental conditioning. All we have shown is that similar principles may underlie both types of learning.

5.4 Temporal Predictions

So far, we have assumed that the context is present simultaneously with reward or punishment. In many cases, it is necessary to include time in a discussion of conditioning. We have already mentioned that in classical conditioning the CS must precede the UC in order for conditioning to take place. This is also true of instrumental conditioning, where the consequences of an action, in this case a reward or a punishment, naturally succeeds the response. To account for this, we need to modify the reinforcement module in figure 5.2.2 slightly.

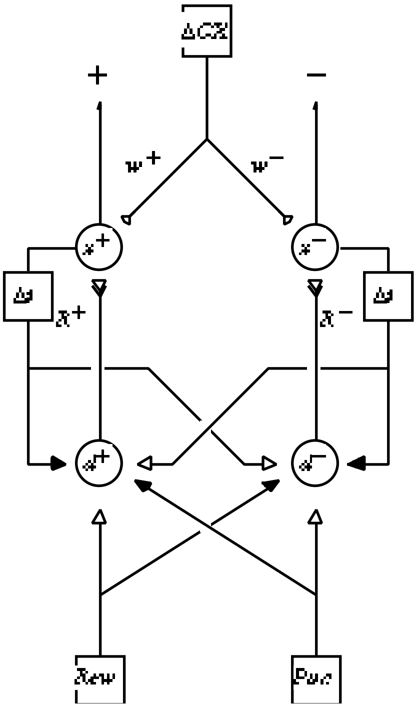


Figure 5.4.1 Reinforcement of temporal predictions. A time–delay is added on the connections from the activating and inhibiting nodes to the reinforcement nodes, which now generates reinforcement only when the CX precedes the reward or punishment.

We introduce a fixed time delay for the connection from the activation and inhibiting nodes to their corresponding difference nodes, and let the learning signal influence the connections, depending on the contextual representation earlier in time (figure 5.4.1). Since it is the time at which each stimulus starts that decides whether one stimulus precedes another or not, it should not be the level of the contextual signals, but rather the level at onset which should control the learning. To indicate this, the contextual input in figure 5.4.1 is called ΔCX. It is assumed that this signal is high when the contextual signal appears and then returns to a low state again. It is, thus, a positive change in the contextual signal which influences learning and not its absolute level (See appendix B and figure 5.4.2). This assumption is made of most contemporary models of conditioning (Klopf 1988, Klopf and Morgan 1990, Mowrer 1960/1973, Sutton and Barto 1990).

Learning progresses in much the same way as in the network shown above except that the context must precede the reward or punishment. There are a number of classical learning experiments that can be explained with this network (figure 5.4.2).

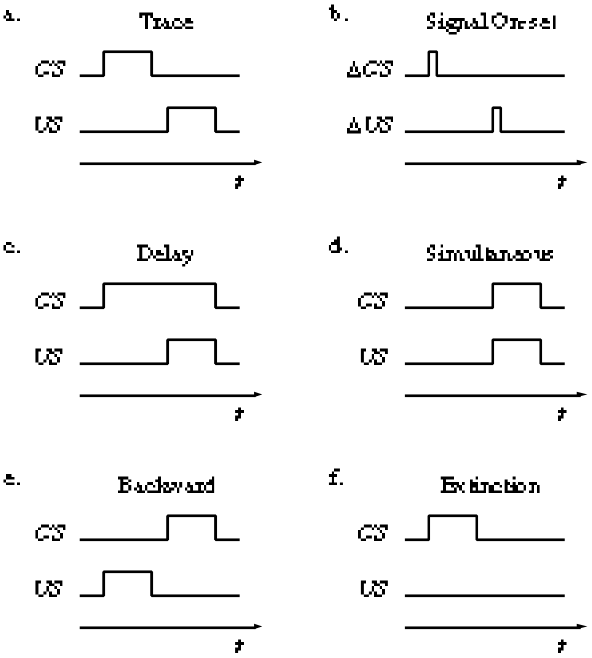


Figure 5.4.2 Classical conditioning paradigms. (a) In trace conditioning, the CS precedes and is terminated before the onset of the US. (b) The onset of the CS and the US in trace conditioning is shown in figure a. Only the onset of stimuli influence learning in the present model. (c) In delay conditioning, the CS is present throughout the presentation of the US. In the present model, this situation is handled identically to trace conditioning. (d) In simultaneous conditioning, the CS and the US are presented at the same time. This does not usually result in any learning. (e) In backward conditioning, the US is presented before the CS. No association is established since the CS has no predictive power in this case. (f) In extinction, the CS is presented on its own. This extinguishes a previously established association with the US.

In *simultaneous conditioning*, the CS and US are presented at the same time. This conditioning paradigm usually results in no conditioning at all (Mackintosh 1983, Pavlov 1927, Smith, Coleman and Gormezano 1969). That is, the US is not able to reinforce any association. This appears to be a rather general property of conditioning, although some counter–examples exist (See Mackintosh 1983)). It is clear that the network presented here will not reinforce a connection if the CS and US are presented simultaneously. This should be contrasted with theories that assumes that connections are established when two signals coincide within the nervous system, such as Hebb's cell–assembly theory (Hebb 1949). The relation between these two views of learning will be discussed in section 7.7.

The learning paradigm that results in the strongest association is *delay conditioning* (Lieberman 1990). In this learning situation, the CS precedes the US and is allowed to be present until the offset of the US. We see that it is necessary to correlate the onset of the signals as described above since the two stimuli are, in fact, present simultaneously. This is the motivation for the use of the change in the contextual signal instead of its absolute level.

Trace conditioning is very similar to delay conditioning except that the CS terminates before the presentation of the US. Since the CS is not perceptible when the US is presented, the conditioning, which usually is quite good, must be the result of a 'memory trace' of the CS, hence the name. Given that the delay between the onset of the CS and the presentation of the US is appropriate, trace conditioning will give the same result as delay conditioning in the present model. According to many animal experiments, trace conditioning usually gives a slightly weaker association than delay conditioning (Lieberman 1990), which is not predicted by the present model.

In *backward conditioning*, the sequence of pairing is reversed. The US is now presented before the CS. As can be expected, no excitatory association is formed in this case, either in the animal experiments, or in the model (Mackintosh 1983). However, some studies have shown that an inhibitory connection is formed, which is not predicted by the present model (See Klopf 1988).

Finally we need to consider *extinction* or *internal inhibition*, which is the process by which an association, or rather, the behavior it produces, is extinguished (Pavlov 1927). An extinction experiment is divided into two phases. In the first, an association is established, and in the second, it is extinguished by presenting the CS alone without the US. In this case, the omission of the US will generate a negative reinforcement signal which increases the weight in the negative side of the system until the inhibiting signal becomes as strong as the activating one.

Figure 5.4.3 shows a computer simulation of the network in figure 5.4.1. In the first phase of the simulation, CS is paired with the US. This will establish an association, w^+ from the CS to the activating node x^+. In the second phase, the CS is presented on its own, and the association will be extinguished again. The extinction is the result of an increase in the association w^+ from CS to x^- that will continue until $w^+ = w^-$. A detailed description of the model can be found in appendix B.

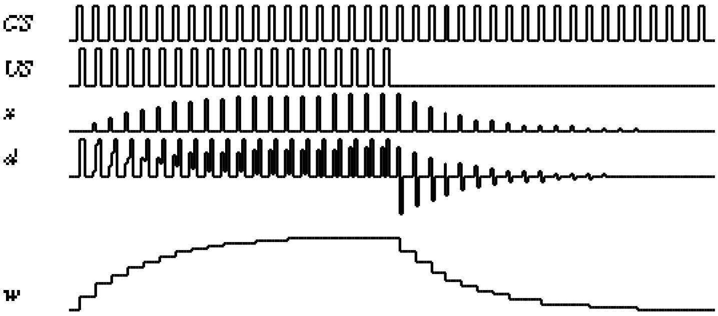


Figure 5.4.3 Computer simulation of conditioning and extinction with the network in figure 5.4.1. Only the sums of the positive and negative sides of the network are shown. For example, the plot named *d* represents $d^+–d^-$.

The learning types above are, of course, only the simplest cases of conditioning. Below we will see what happens when many stimuli are allowed to interact with each other.

5.5 Multiple Cues

To make the conditioning situation more realistic, we need to consider what happens when many contextual cues are present simultaneously. The most important situation where many cues interact is in a *blocking* experiment (Kamin 1968). In this type of learing, CS_1 is first paired with the US using any of the paradigms described above until an association has been established. On later trials, two cues, CS_1 and CS_2, are simultaneously paired with the US (figure 5.5.1). Had the CS_2 been paired on its own, it would have established an association with the US, but when it is paired together with the CS_1, the result is different. If the CS_2 is later presented on its own, no CR will be produced, which means that no association has been formed between the CS_2 and the US (or UR).

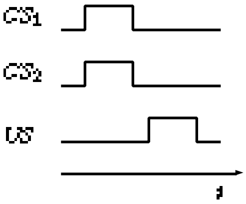


Figure 5.5.1 Multiples cues. In a blocking experiment, an already established association between CS_1 and the US, blocks the formation of a new association between CS_2 and the US. If the two cues have not been paired with the US before, they each learns an association with half the strength compared to if only one cue was present.

The explanation of this phenomenon is that the CS_1 is sufficient to predict the US and therefore blocks the CS_2. This is a natural consequence of the fact that it is the difference between the expected reward (or US) and the actual reward which reinforces the learning process. Since the US is already predicted, no reinforcement will be generated. The present model has this is property in common with both the Rescorla–Wagner model (Rescorla and Wagner 1972) and various neural-network models using the delta-rule and its variations (Widrow and Hoff 1960).

Figure 5.5.2 shows a computer simulation of a blocking experiment using the network in figure 5.4.1. First the CS_1 alone is paired with the US which will set up an association from CS_1. In the next phase of the experiment, the compound $CS_1 \bullet CS_2$ is paired with the US. Since the CS_1 sufficiently predicts the US, no association from CS_2 will be formed. The association from CS_1 will, thus, block the learning of the association from CS_2.

A consequence of the way the reinforcement signal is computed is that the learning system is insensitive to the number of contextual cues present. If two cues CS_1 and CS_2 are paired with the US, they will each receive half the associative strength that would otherwise be given the single cue. If more than two cues are present, the prediction of the US will still be at the same level.

An important consequence of this property is that the CS need not be a single signal. It can also be a whole perceptual schema represented over a large number of nodes (See Balkenius 1992, 1994c and chapter 9). Thus, the perceptual system need not know about the learning system. Whatever representation it produces, the learning system will be able to use it as long as it uses a spatial code. Representations of this type are sometimes called *representation by place* or *labeled-line coding* (Martin 1991). The essential property of this type of representation is that the same signal always codes for the same perceptual property[2].

A situation related to blocking is *overshadowing*. In this type of experiment, two cues are simultaneously presented, but one has greater salience than the other. In the present model, this is represented by a larger signal for the more salient cue. As a consequence, this cue will overshadow the other and receive most of the association. This property follows from the fact that the strength of the contextual representation influences the learning speed (See appendix A).

Let us now consider a situation were two cues, CS_1 and CS_2 have been paired with the US at different times in such a way that they both have established an association. The presentation of either CS_1 or CS_2 on its own will cause the CR to be

---

[2] Contrast this type of representation with the type used in a computer or in a telephone line where the momentary signal, in itself, does not mean anything, unless it is related to other signals that are sent on other lines or at other times. Labeled-line coding is more similar to the representation in the wire from the button on the door to the door bell. This signal means that the button is pressed, no matter in which context it is received. No decoding is necessary.

produced. We can arrange for two types of relations between the compound stimulus representation of CS_1 and CS_2 which we will call $CS_1 \bullet CS_2$. Either the compound predicts US, as well as the individual cues, or it does not (figure 5.5.3).

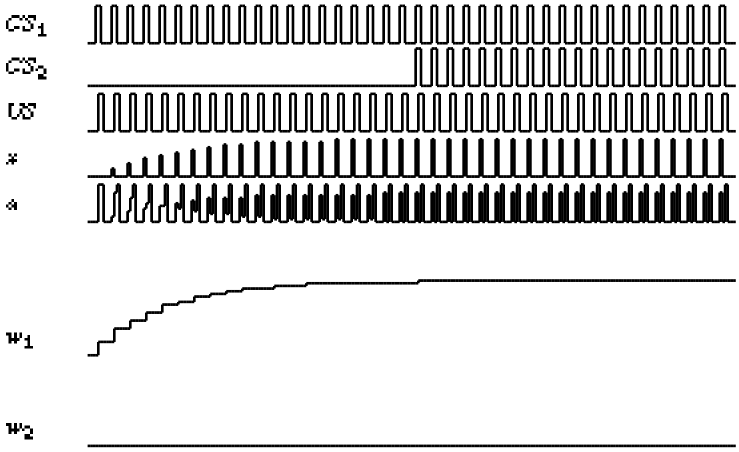


Figure 5.5.2 Blocking. In the first phase of the simulation, an association from the CS_1 is established through the connection w_1. In the second phase, the compound $CS_1 \bullet CS_2$ is paired with the US. Since the US is already entirely predicted by CS_1, the association with CS_2 will be blocked, that is, the weight on connection w_2 will not increase.

a. $CS_1 \rightarrow US$
$CS_2 \rightarrow US$
$CS_1 \bullet CS_2 \rightarrow US$

b. $CS_1 \rightarrow US$
$CS_2 \rightarrow US$
$CS_1 \bullet CS_2 \rightarrow$ no US

Figure 5.5.3 Compound predictions. (a) CS_1, CS_2, *and* their compound predict the US. (b) Negative patterning. Each CS on its own predicts the US, but the compound does not.

The first case in figure 5.5.3a is, of course, the usual relation between the compound and the US. Since each cue predicts the US on its own, the compound prediction will be an US of twice the strength as only one cue. This will cause the CR to be produced while simultaneously acting as a case of partial omission, since the US was only at half the expected level. Still, positive patterning is easily handled by the reinforcement module presented above.

The second situation is called *negative patterning* (Kehoe 1990, Roitblat 1994). In this case, the compound is not followed by the US. In the neural network literature, this situation is known as the XOR problem, since the logical function from the cues to the US is “exclusive or”. It is well known that this problem cannot be solved with a single layer of nodes (Minsky and Papert 1988). However, a number of network architectures exist that are able to solve it. The main idea behind these models is to generate a category node which codes for the compound and let the signal from this node override the signals from the individual cues. Since the process of categorization is related to perception, it will not be considered until chapter 7. Below we will see what happens if we introduce the cues at different times instead of simultaneously.

5.6 Higher–Order Conditioning

Let us assume that CS_1 has been paired with the US until an association has been established. What will happen if another cue, CS_2, is now paired with CS_1? It turns out, quite sensibly, that the CS_2 is now able to produce the CR, that is, the CS_1 is able to reinforce an association from CS_2. This is called *secondary* or *second-order conditioning* and shows that it is possible for an initially neutral stimulus to act as an US for a secondary learning process once it has been paired with the US. In an instrumental learning situation, the CS_1 will act as a secondary reward.

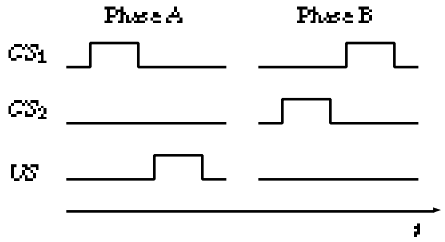


Figure 5.6.1 Second-order conditioning. In phase A, the CS_1 is paired with US until an association has been established. In phase B, the CS_1 is used to reinforce an association from CS_2.

There also exists another variation on second-order conditioning. This is the case when CS_2 is first paired with CS_1, and only at a later time is the CS_1 paired with the US. This procedure is also able to make CS_2 produce a CR in some cases. In this section, we will only consider the first type of secondary conditioning. The other type will be discussed in section 5.8 below.

In figure 5.6.2, the reinforcement module has been extended with direct connections from the activation and inhibition nodes to the reinforcement nodes (d^+ and d^-). Since these direct connections are not time–delayed, the signal received at the activation node at time *t* will simultaneously act as a reward at the reinforcement nodes. The activity at the inhibition node will act as punishment in a similar way. Since the expected reward and punishment is delayed on its way to the delta nodes, it will not interfere with the secondary learning process. In this way, the contextual signals can act as reward or punishment at time *t* and as expected reward or punishment at time *t*+1. This network is, thus, able to model primary as well as higher-order conditioning.

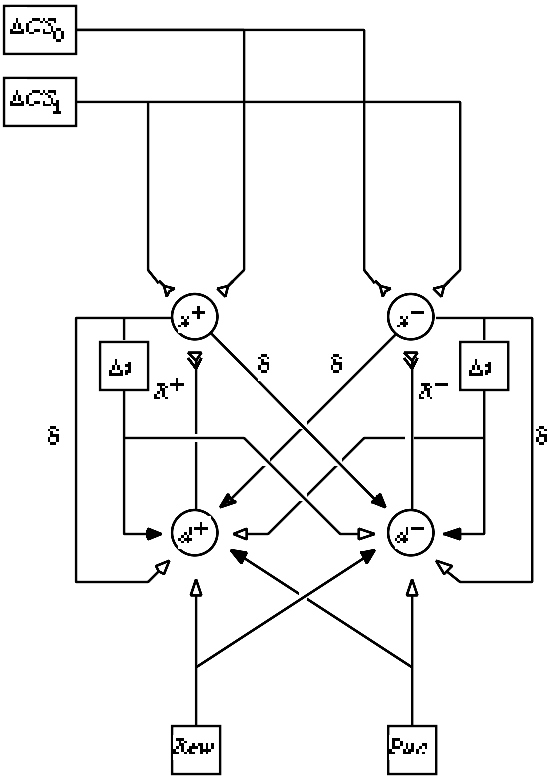


Figure 5.6.2 Higher-order reinforcement. The input to the activation and inhibition nodes at time *t* are used for higher-order reinforcement at time *t* and as expected reward or punishment at time *t*+1. Note that the network is almost identical to the one in figure 5.4.1.

Figure 5.6.3 shows a simulation of secondary conditioning with the network in figure 5.6.2. As can be seen, the pairing of CS_1 with the US increases the weight w_1^+. In the second phase, CS_2 is paired with CS_1 which will now generate secondary reinforcement. This will increase the weight w_2^+. Since the US is not presented in this phase, the CS_1 will start to extinguish by increaseing the weight w_1^-. As a consequence, the association from CS_2 will also extinguish later on.

The transient nature of secondary conditioning has led some researchers to believe this the effect is not real (See discussions in Rescorla 1980 and Klopf 1988). The model presented here predicts that secondary conditioning should behave in this way, however. It also suggests that a strong association can be formed by secondary conditioning if CS_1 is still followed by the US in the secondary learning phase. This would still be a case of secondary conditioning as the US is not able to reinforce the connection from CS_2 because it occurs too early in time.

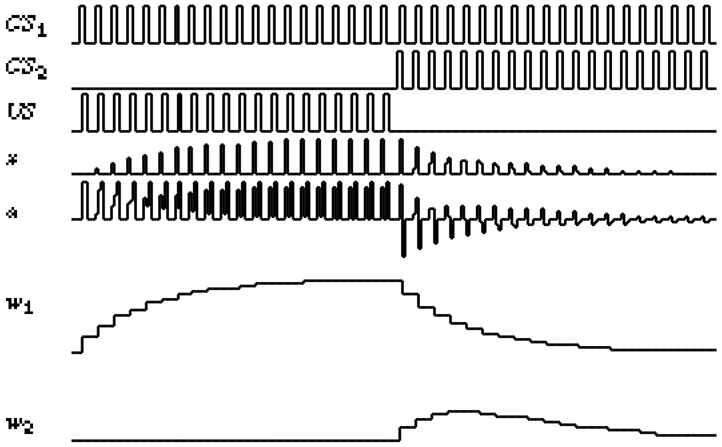


Figure 5.6.3 A simulation of secondary conditioning. Only the sum of the activation and the inhibiting connections are shown. As can be seen, the secondary association between CS_2 and the US is only transient since the US does not follow CS_1 in the second phase.

Figure 5.6.4 shows a computer simulation of such a learning experiment. In this case, the second-order association is not transient. This makes it possible to use CS_2 to establish further associations. In section 7.9, this property of the reinforcement module will be used for sequential learning.

As sequences of stimuli are presented to the network, long sequences of predictions can be formed when stimulus CS_3 is reinforced by CS_2, and CS_4 is reinforced by CS_3, and so on. Should this process continue indefinitely, all stimuli would eventually produce the CR. Since this is not generally desirable, we introduce a weight less than 1 on the connections that mediate the secondary conditioning. This weight, δ, will be called the *discount factor* for the secondary conditioning. For instrumental learning it corresponds to the idea that a reward now is worth more than a reward at a later time.

The discount factor, thus, plays a role similar to interest rate in economics. It determines to what extent a reward can be postponed until a later time. With a high discount factor, the value of a reward does not change much with time. With a low discount rate, it is more preferable to receive a small reward now than a larger one later on.

If the US occurs at time t_{US}, the discounted prediction from a CS at time t_{CS} will be,

$$\delta^{(t_{US}-t_{CS})}US \tag{5.6.3}$$

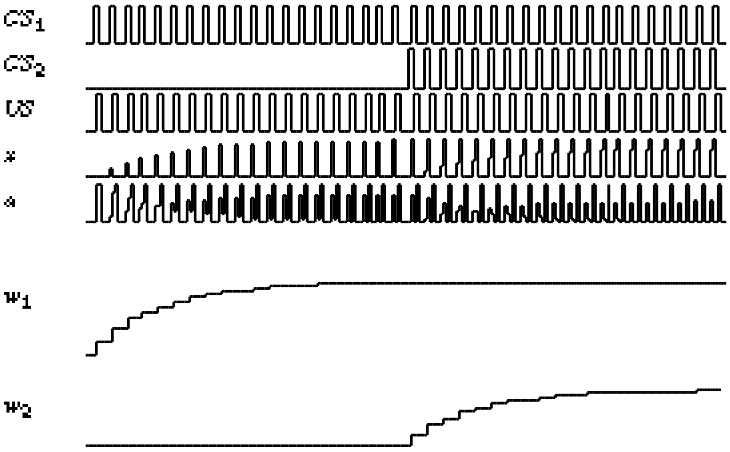


Figure 5.6.4 A simulation of modified secondary conditioning. The secondary association between CS_2 and the US grows to a high level when the US follows the CS_1 in the second phase. A discount factor of 0.9 was used in the simulation which means that the association from CS_2 will approach 0.9 times that of CS_1.

This means that the CR will be weaker, or less probable, the further away in time it occurs with respect to the US. Also note that it is necessary for the reinforcement module to predict not only that the US will occur, but also the *level* of the US. We can, thus, conclude that conditioning should be driven by the time relation between the onset of the CS and the US or secondary reinforcing CS, and should try to make the net activity at the activation and inhibiting nodes approach the discounted level of the US. See appendix A and B for a more exact description of the reinforcement module.

We can now return to the discussion of psychological distance we started in section 4.2. There is an obvious similarity between the idea of a discounted reward and a potential reward function. Let R_g be the reward obtained at the goal location g, and let $c(z,g)$ be the cost of performing the actions that lead from z to g. The discounted reward, $D(z,g)$, at location z, with respect to the goal at g, can, thus, be defined as,

$$D(z,g) = \delta^{c(z,g)} R_g \qquad (5.6.4)$$

This is, therefore, an appropriate internal estimate of the cost of moving from z to g, which is represented as x^+ and x^- when z is the current context. If $CX(z)$ is the current contextual input, the net activity, $x^+ - x^-$ should try to approach $D(z,g)$. The relation between the discounted reward and the potential reward function is given by,

$$\delta^{c(z,g)} = \frac{\delta^{-G(z)}}{\delta^{-G(g)}} \qquad (5.6.5)$$

Here, of course, the discount factor does not depend on time, but on cost. When time is the only relevant variable to optimize, the two measures coincide. When it is not, it would be fruitful to redefine conditioning in terms of cost instead of time. This would make it possible to use instrumental conditioning to learn the optimal behavior with respect to a cost function and not only to time. We will not pursue this task here, however.

The potential reward function is a measure that does not depend on the animal while the discounted reward function, $D(z,g)$, is a similar measure based on what the animal has learned. It can, thus, be used as a basis for a definition of psychological distance.

Let us define psychological distance as the estimated cost of moving from a state z to a goal state g based on the discounted reward as,

$$\psi(z,g) = \frac{\ln(D(z,g)\ /R_g)}{\ln \delta} \tag{5.6.6}$$

In the special case when only time is optimized, the psychological distance is simply the time difference between the occurrence of a CS, and the occurrence of the US, thus,

$$\psi_{TIME}(CS,US) = t_{US} - t_{CS} \tag{5.6.7}$$

In either case, the psychological distance between two arbitrary situations x and y, with respect to a goal g, can be calculated as,

$$\psi_g(x,y) = \left|\psi(x,g) - \psi(y,g)\right| \tag{5.6.8}$$

The inclusion of the goal in this calculation is, of course, not very satisfactory, but is necessary, since all learning is made with respect to a goal reward or an US. In section 5.8, we will see how this requirement can be relaxed in a general form of expectancy learning, but first we need to consider how the learning mechanisms above are used in different engagement systems.

5.7 Appetitive Learning

In chapter 4, we saw that only a fairly simple nervous system is necessary to make a creature approach a goal object in a stable way. However, such an ability depends in a critical way on three factors. The first is that the animal has innate knowledge about which objects, and, thus, which smells, that are appetitive. The second requirement is that the creature has an adequate sensory system to determine the correct distance to the goal. Finally, it is necessary that the animal can sense the goal from its initial position. In this section, we will take a look at a number of learning mechanisms that can be used to relax the first two requirements. The last limitation will be handled in chapter 8.

Learning How to Approach

As we saw in section 5.3, a creature using a combined approach strategy must have a correct estimation of the distance to the goal if it wants to approach it successfully. When the estimated distance to the goal is to low, it will stop before the goal is reached, and if the estimation is too high, the creature will either run past the goal or get hurt while colliding with it. To behave optimally, the creature should stop exactly at the goal, that is, the approach and avoidance gradients should be equal at the goal location (See figure 4.2.13).

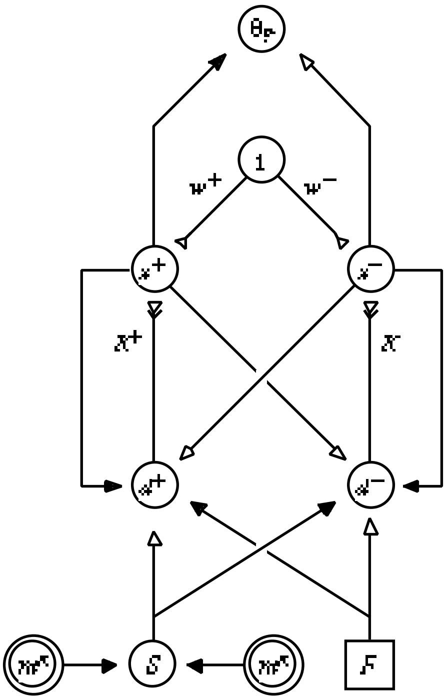


Figure 5.7.1 Using the reinforcement module to change the threshold for the passive avoidance gradient. The threshold increases when the creature stops without sensing food, or when food is sensed without the creature stopping. The node *S* detects that the creature has stopped and acts as a 'reward', that is, it increases the passive avoidance threshold. The food detector, *F*, acts as a punishment and decreases the threshold. When the creature is moving without sensing food, or has stopped at the food, the two signals cancel each other and no learning takes place.

There are obviously two situations were an incorrect gradient can be detected. In the first case, the creature stops as a consequence of the two gradients being equal, but does not receive any reward. In this case, the avoidance gradient is too high and should decrease. In the second case, the animal senses the goal, but does not stop. In this case, the avoidance gradient is too low and should increase. In both cases, the situation can be corrected by changing the threshold for the avoidance gradient (See equation 4.2.8). This is equal to changing the weight on a connection from a node whose output is always one to the node θ_p in figure 4.2.10. In figure 5.7.1, a network is shown that can change the passive avoidance gradient in this way.

Learning What to Approach

It is of great importance for an animal to be critical about what it eats. Before a piece of potential food can be ingested, it must be thoroughly examined to determine whether to consume or to reject it. The primary modalities involved in this process are taste and smell. These two sensory systems complement each other in a number of useful ways.

The olfactory system reacts at a fairly long distance and on a very large set of stimuli. Although the exact mechanisms used in the olfactory receptors are not known, it is clear that the number of distinct smell sensations is almost unlimited (Davis and Eichenbaum 1991). These can be organized into three classes: appetitive, aversive and neutral. As presented in the previous chapter, appetitive and aversive stimuli are defined as stimuli which the animal will approach and avoid respectively. All other stimuli are neutral. As far as we know, it has not been established to what extent appetitive smells are innately known, but it is well known that many smells, such as hydrogen sulfide, are inherently aversive. The situation is different for taste, however. There are four, or possibly five (Kandel, Schwartz and Jessel 1991) basic types of tastes that can be combined to produce complex sensations, but they also have distinct meanings on their own that are directly relevant to the animal.

A *sweet* taste indicates that the food has a high level of energy. This is of obvious importance to an animal, and we would expect these taste sensors to play an important role in learning. This is indeed the case. Since biochemical processes are highly sensitive to the acidity of the environment in which they occur (Tortora 1990), it is not very surprising that one set of taste sensors are devoted to the measurement of the acidity of a potential food object. This is the role of the sensors for *sour* taste. Another factor of great importance for the body cells is the concentrations of sodium and chlorine ions. Chlorine is important for the water movement between cells, and sodium is necessary to maintain the water balance in the blood (Tortora 1990). Not surprisingly, one type of smell receptors has been assigned to the detection of *salt*.[3] The final type of taste, *bitter*, differs in two important ways from the other three. While the tastes of sweet, salty and sour can be easily related

[3] We do not mean to imply that the behavioral mechanisms described here are the prime mechanisms for maintaining homeostasis. Although it is important to keep, for example, the level of sodium at an appropriate level, this is handled primarily by internal processes, and not by ingesting more or less salt.

to various chemical substances, this is not the case for bitter. There are, of course, well known substances that will make the these receptors react, but there does not seem to be a clear class of substances that have a bitter taste. The second difference is that a bitter taste is generally aversive. Taste sensors are, thus, the prime candidates for primary reward and punishment signals in the food–related appetitive engagement module.

An appropriate strategy for an animal would, then, be to approach and consume objects that have a sweet taste and avoid, or at least, ignore objects that taste bitter; but since the taste receptors only react when the food is in the mouth, it is necessary to use some other modality to localize the food. When food is eventually found, the animal can taste it and determine whether to consume or reject it. The simplest behavioral strategy that could combine taste and smell would consist of, at least, three phases: (1) Approach any appetitive smell; (2) Taste the potential food object; (3) Consume or reject based on taste.

It is quite possible that an animal could survive using this strategy, but it is not very productive if the environment consists of many objects with appetitive smell, but no nutritious value. An unnecessary amount of time would be dedicated to the approach of useless objects. The obvious solution to this problem is to let the animal remember whether a certain smell predicts an appetitive or aversive taste, that is, by including a learning process (Lynch and Granger 1991).

In section 2.10, we made a distinction between three different types of learning called early, synchronous and late, depending on where it occurred with respect to the consummatory situation. In learning what to approach, or more specifically, what to eat, all three types of learning are useful. The association between smell and taste can be seen as an example of synchronous learning. Taste is only available in the consummatory situation, and this is, thus, the only time when this type of learning can occur. The illness caused by poisonous food naturally occurs after the consummatory situation and is, thus, an example of late learning (See section 2.10). As we will see below, early learning plays an important role when food is not immediately accessible, but can be found only after a large sequence of behaviors.

Figure 5.7.2 shows how the reinforcement module can be connected to a behavior module for approach and consummation in such a way that the taste of food determines whether it will be approached again or not. The behavior module to the left in the figure implements one of the approach behaviors presented in section 4.2. The output from the behavior module is facilitated by the activating output from the reinforcement system. The inhibiting output controls the threshold of the passive avoidance gradient through the node θ_p. The signals from the smell receptors are used both to control the approach behavior and as contextual input to the learning system. A sensor for appetitive taste, T^+, is used both as reward and as a signal that starts the eating behavior. The eating behavior will also temporarily inhibit the motor neurons to let the creature stay at the food while eating. The network presented in the previous section could alternatively have been used here instead. This is, thus, a simple case of an appetence and consummation hierarchy.

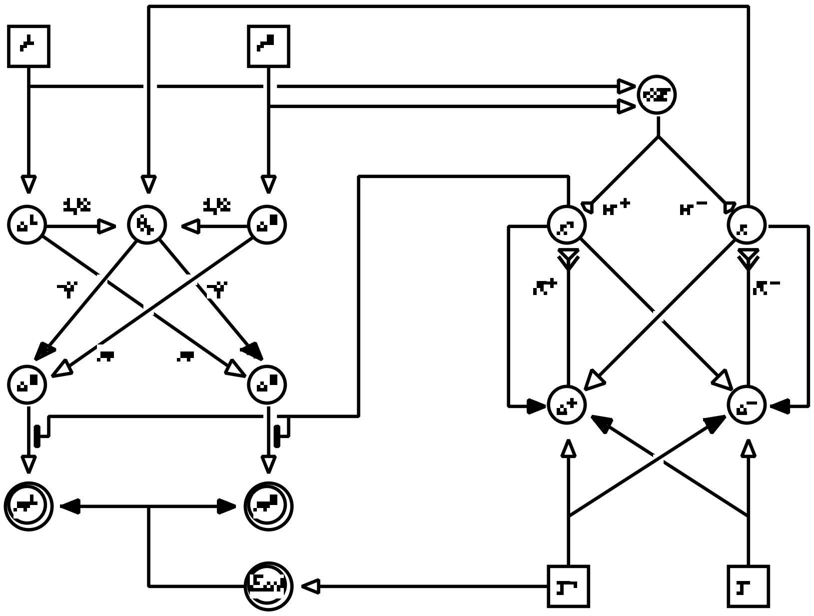


Figure 5.7.2 A simple engagement system for appetitive behavior. The network to the left implements an approach behavior with potential behavioral inhibition. The network to the right controls whether a specific smell should be approached or not. The connection weights must be set as shown in figure 4.2.11. Note that the signals from the olfactory sensors act both as controlling stimuli and as context for the reinforcement module.

Learning not to Eat

With the nervous system in figure 5.7.2, the creature will successfully learn which smells predict palative food, given that the taste reflects whether the food object is nutritous or not. There are also cases in which a food object does indeed taste good, but still should not be immediately consumed. This is the case with novel foods that could potentially be harmful although they do not taste that way.

There are a number of things we must take into consideration for a system of this kind. To start with, is the food known? If it is, prior knowledge of the object can be used to determine whether it should be eaten or not, as we saw above. If the object is new, it should be tasted, but with caution. Only a small amount should be consumed at first, and if the creature subsequently becomes sick, it should remember the food as aversive.

This strategy requires, first of all, that the creature determines if a piece of food is unknown. We have already seen that the reinforcement network can compute novelty in the required way. We simply use the output from the delta nodes and compare it with a fixed treshold to determine whether a smell is known or not. If it is not, a special eating behavior is used that will make the creature consume a small amount of the novel food and remember its smell if it later becomes sick.

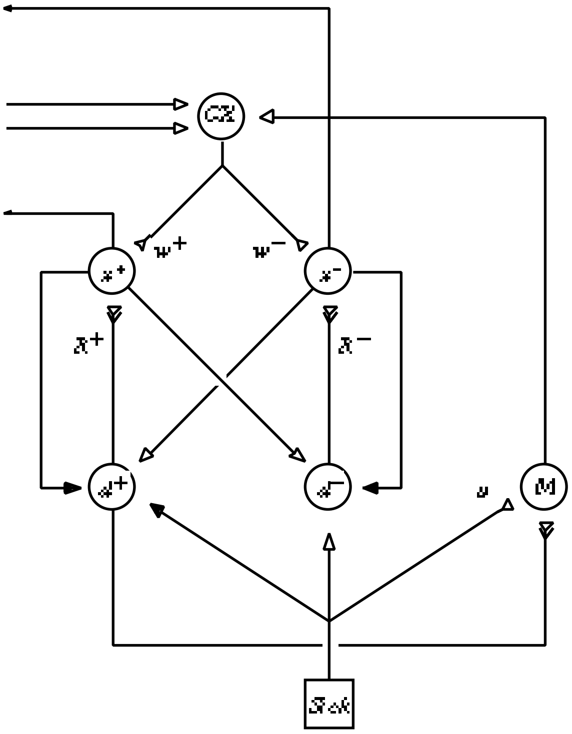


Figure 5.7.3 Smell aversion based on the network in figure 5.7.2. The additional circuitry needed for a working memory of novel foods. If the creature becomes sick, the smell of the new food will be recalled at *CX* and the reinforcement system will learn to inhibit approach toward that smell.

Figure 5.7.3 shows an extension of the previous network which includes a rudimentary working memory in the form of a plastic connection, *u*, between the illness sensor, *sick*, and a memory node *M*. When the output from d^+ is positive, the network has not yet learned the reward for the current food, and it should, thus, be remembered as new. This signal is used to temporarily raise the efficiency of the plastic synapse, *u*. Note that we need not store a representation of a food that tastes bad since it will get immediately rejected. The output from d^- can, hence, be ignored here. If the creature becomes sick while the synapse *u* has raised efficiency, it will recall the smell of the unknown food by reactivating *CX* through the node *M*. The punishment, from *sick*, will, thus, not act on the current sensory input, but on a stored representation of the novel food.

Note that in a larger system, there need be one memory node for each contextual input. It is possible to move the learning from the connection *u* to the pathway from *M* to *CX*. This would make it possible for the memory node to reactivate a contextual pattern instead of a single node. This would require fewer nodes and would, thus, be more economical in a larger nervous system (See section 9.4).

In real animals, a mechanism similar to the one described above shows up in the form of *taste* aversion, and not smell aversion. In chapter 2, we saw that rats very easily learn to associate taste with sickness (Garcia and Koelling 1966, Garcia, McGowan and Green 1972). It is not unlikely that a mechanism similar to that in figure 5.7.3 is involved. This network also introduces two very important ideas.

The first is that the learning signal generated in the d^+ node is stored in a working memory. The output from d^+ can be considered to signal novelty and the reinforcement circuit can be seen as a novelty filter (Kohonen 1989, 1984). There exists neurophysiological evidence for the existence of a mechanism of this kind in the subiculum of the hippocampal formation (Shepherd 1990, Gray 1982). It is also well known that this structure is also involved in the temporary storage of memories (Brown 1990, Eichenbaum *et al*. 1991, Mishkin and Petri 1986, Olton, Branch and Best 1978, Squire 1992).

The second mechanism of importance is the recall mechanism. When a specific cue is presented, in this case a signal from the sickness sensor, a specific ‘memory’ is regenerated at *CX* and can be associated with events which happened long after its initial activation. Extensions of this mechanism that can handle more complex memories will be discussed below. For example, the solution to the radial maze depends to a high degree on a learning mechanism of the type that can recall the previous choices of arms (See section 2.8). Such an extension depends on an ability to recall memories using arbitrary cues and not only with a sickness signal, as above.

5.8 Aversive Learning

The reinforcement module presented above can also be used for escape learning. For this engagement system there exists no rewarding stimuli, and all learning has to be controlled by the termination, or omission, of some aversive stimulation. The prime aversive state is, of course, the presence of a predator on hunt.

Since the aversive state will in itself cause the creature to flee, one may wonder whether it has anything to gain from an aversive learning ability. If the only thing the creature will learn is to escape, which it would do anyway, there appears to be no need for this learning mechanism. If we assume, on the other hand, that there exist certain variations in the way the creature escapes or avoids the predator, learning could in fact be useful.

To use the reinforcement module for escape and avoidance learning, the termination or omission of punishment, that is, the aversive state, should reinforce the behavior that preceded it. For real animals, the situation is usually much more complicated, however. Remember that real animals have species-specific defense mechanisms that are performed in aversive situations. These behaviors will usually interfere with aversive learning in various complex ways (See section 2.4).

5.9 Learning Action Sequences

In many cases, a desired goal cannot be approached in a single step. It may be necessary to perform a whole sequence of actions before the goal can be reached. In section 4.3, we saw how behavior modules can be linked together in sequences in a number of ways. In this section, we will see how secondary conditioning can chain behaviors together by building an internal estimation of the potential reward gradient.

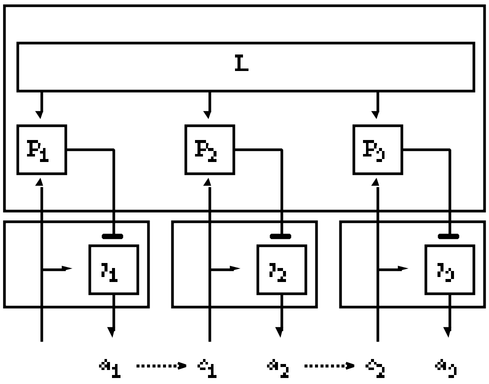


Figure 5.9.1 The general structure of a behavior chaining mechanism. The applicability predicates of the behavior modules are moved into a learning system, L, that determines which behavior to produce based on the history of primary and secondary reinforcement. The arbitration of the different behaviors can be handled by any of the arbitration mechanisms described in chapter 4. Any of the linking schemes described in section 4.3 are possible, although only external linking is shown in the figure.

In section 2.7, four types of chains were presented. The first and simplest consists of a stimulus which by itself produces a whole sequence of responses. The second type of sequence consists of a set of S–R associations where each response generates the stimulus that triggers the next S–R association. As described in section 4.3, this linking can be either internal to the organisms or rest on response dependent changes to some observable property of the external world. The third type of sequential structures consists of stimulus–approach associations, where the behavior generated by the animal is guided by an external goal stimulus. Finally, we considered a chain of place–approach structures, where each step in a sequence is guided by a whole set of external stimuli.

Each of the four types of behavior sequences can be considered as an instance of the general layout illustrated in figure 5.9.1. The applicability predicate of each behavior module is moved into a learning system that determines which behavior should be produced, based on the history of reinforcement. Depending on the type of behavior modules involved, the architecture can generate any of the four types of chains described above.

There are two fundamental ways in which chaining of behaviors can be accomplished. In the first case, only primary reinforcement is used. All behaviors that have occurred prior to the reward are strenghened, but the size of the reinforcement depends on when the behavior occurred relative to the reward. The behavior executed immediately before the reward is stregthened most, and the further away a behavior is in relation to the reward, the less it is strengthened. Another type of chaining mechanism depends on *conditioned* or *secondary reward* or *secondary punishment*.

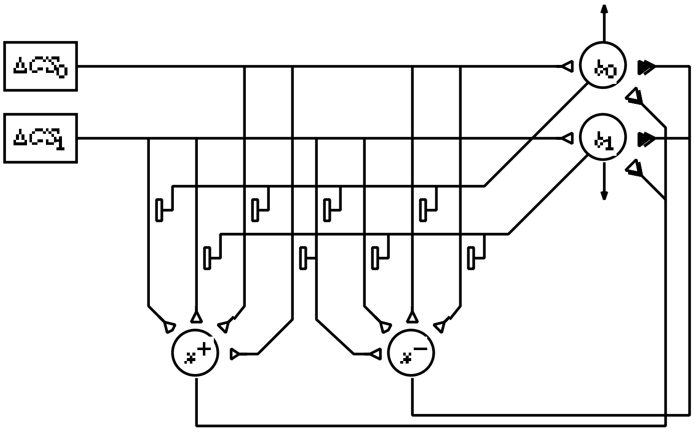


Figure 5.9.2 Learning of behavior sequences. Two sensory inputs are associated with the behavior that predicts the highest reward. The two nodes b_0 and b_1 activate different behavior modules and the nodes x^+ and x^- are part of the network shown in figure 5.6.2.

Figure 5.9.2 shows how sensory cues can be associated with behavior modules using the network for secondary conditioning described above. There is one connection from each sensory input to both the activating and inhibiting nodes of the reinforcement module for each behavior module in the system.

The connections from the sensors to the reinforcement system are responsible for the learning of the potential reward in each situation. This potential reward is used to reinforce the connections from the sensors to the behavior modules. The learning in these connections tries to make the weights approach the expected reward which is received if the behavior is performed. It is necessary that some form of arbitration is used to select which behavior should be active at the times discussed in section 4.3. The architecture of this part of the network is very similar to that used by Klopf, Morgan and Weaver (1993), although the dynamics of the reinforcement network is different.

We will return to this network in chapter 8 were it will be one of the basic learning mechanisms for spatial orientation. A formal description of the network can be found in appendix D.

5.10 Expectancies

We are now finally ready to introduce the most powerful learning mechanism we will consider in this book: learning of expectations of stimuli. The networks presented above were all involved with the learning of an expected reward, punishment or an US. However, it is also possible to design a network which learn S–S expectancies. Such a network is very different from the ones above in that the associations learned initially have no specific meaning. This makes the structures learned more flexible because they can be used in many different contexts. In later chapters, we will see how learning of this type can be used in a number of systems, from habituation of exploratory behavior to choice behavior and planning. These are very complex systems, but the basic circuit is an old friend, the reinforcement module.

In figure 5.10.1, the basic structure of the expectancy network is illustrated. Each sensory input has its own reinforcement module that can change the associations with its corresponding stimulus. If CS_1 is followed by CS_2, a connection will be established between CS_1 and CS_2. The connection will be reinforced by CS_2, as if it acted as secondary reward for its reinforcement module.

Since the same learning is used here as for the expectation of reward above, all properties of classical and instrumental learning carry over to this network, the only difference being, that all stimuli are able to reinforce all others. There are no stimuli that have precedence over the others. This is, thus, an example of early learning (See section 2.10). This leads naturally to the view of learning as a process where animals learn about causal relationships between events (Dickinson 1980).

An interesting consequence of the learning in this network is that the network will read out its expectations. If a sequence has been presented to the network repeatedly, such as CS_1, CS_2, CS_3, the presentation of CS_1 will automatically regenerate that sequence in the network. At each step, the anticipated stimulus will be compared with the actual input at that stage, and the delta nodes will signal if something unexpected occurs. If expectations are not met, they will gradually become extinguished.

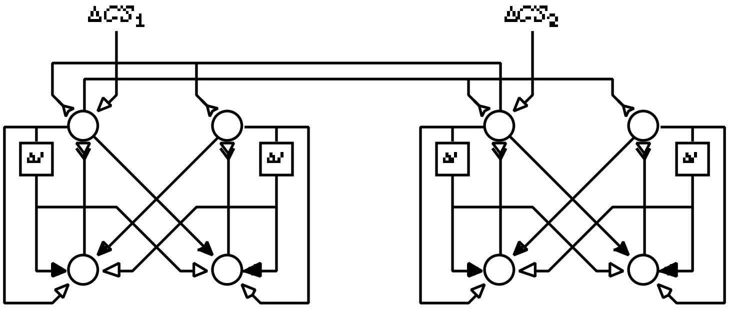


Figure 5.10.1 Stimulus–stimulus association between a set of reinforcement modules. Each input can reinforce associations to itself without any need for primary reward.

One problem with this network is that it uses a fixed time delay in the conditioning process. This means that expectancies cannot be learned over other time delays. A simple way to overcome this problem is to let each sensory input trigger an avalanche, as described in section 4.4. Such a representation is sometimes called a multiple-element stimulus trace (Desmond 1990). Each node in the avalanche will code for a specific time–delay after the onset of a stimulus. This makes sure that associations can be formed over arbitrary time delays.

We can now define psychological distance without reference to a reward as follows. The psychological distance between *a* and *b* is given by,

$$\psi(a,b) = \frac{\ln D(a,b)}{\ln \delta}\,, \tag{5.10.1}$$

where, $D(a, b)$ is the discounted prediction of *b* given *a*, as in equation 3.6.2. In the case when only time is optimized, the psychological distance is the time difference between the occurence of a CS_1, and the occurence of the CS_2, thus,

$$\psi_{TIME}(CS_0, CS_1) = t_{CS_1} - t_{CS_0} \tag{5.10.2}$$

Given these properties, we see that a combination of the S–S network and the S–R network above can explain secondary conditioning in the case where the pairing of CS_2 and CS_1 precedes the pairing of CS_1 and the US. Since CS_1 can reinforce the association from CS_2 to CS_1 even before CS_1 has been paired with the US, the order in which the individual pairs are presented is irrelevant.

We will return to this network in chapter 8 and 9, where we will discuss its role in complex cognitive processes where it is used to anticipate future percepts to guide behavior by prediction of the future rather than the immediate perceptual environment. A formal description is given in appendix E.

5.11 Limitations of the Reinforcement Module

There are a number of important limitations of the reinforcement module described in this chapter. The first is that it almost totally ignores interstimulus interval (ISI) effects. In the model, conditioning is only possible for a fixed time interval, and during this interval, the effect is the same regardless of the precise timing of the stimuli. Figure 5.11.1 shows the approximate relation between learning and interstimulus interval found in empirical studies (Mackintosh 1983). The shape of this curve is approximately the same for all species, although the time scale can vary substantially. For example, for rabbit eyelid conditioning, the curve describes approximately a second (Smith, Coleman and Gormezano 1969), but for pigeon key-pecking, the same curve covers around half a minute (Gibbon *et al*. 1977). For food aversion, the ISI may be as much as 24 hours (Mackintosh 1983).

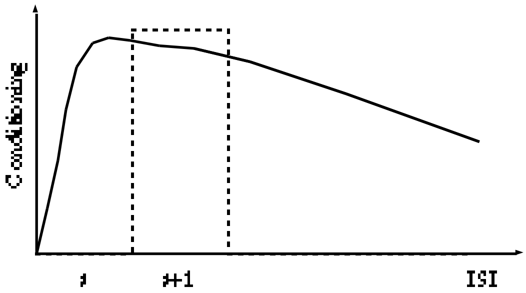


Figure 5.11.1 The effect of interstimulus interval on conditioning. The solid line shows the approximate relation between interstimulus interval and strength of conditioning found in empirical studies. The dotted line shows the behavior of the present model.

The model can easily be extended to handle ISI effects. Such an extension is described in appendix C, but will not be discussed here since it would make the presentation unnecessarily complicated.

A second limitation is that reinforcement is only generated at CS onset. It has been found in empirical studies that CS offset also triggers learning, but of opposite sign (See Sutton and Barto 1990). To model this phenomenon within the present model, it would be required that the ΔCS be negative as well as positive. The only reason why this mechanism was not included above is that it requires the connections to the activation and inhibiting nodes to be negative as well as positive, which is unrealistic from a neural point of view.

Another notable omission is that we have not considered *learned irrelevance* or *latent inhibition* (Lubow and Moore 1959). This is the situation when an animal learn that a stimulus does not have any consequences at all and is, thereafter, reluctant to use it as a cue. Like many related models, (for example Klopf 1988), the present model fails to handle this situation unless a background stimulus and an internal trial clock is postulated. In chapter 9, we will return to this type of learning and suggest that it may depend on cognitive processes of a very high level. A related phenomenon is *overshadowing* where a stimulus compound is paired with an US, but some parts of the compound do not establish any association. The part of the compound that does establish a connection is said to overshadow the other parts. Although the present model does model overshadowing, the way it does it is not entirely satisfactory. An alternative explanation that sometimes seems more fruitful is that an attentional process selects a single stimulus when many are present, which would also result in overshadowing (Mackintosh 1974). We will return to this alternative in chapter 9.

A final simplification is that we have assumed that the reinforcement module is composed of two identical halves. The system for reward is identical to the system for punishment. There is, however, a large difference in the way reward and punishment affect behavior. When an animal is rewarded, it is likely to perform the rewarded behavior again. This will let the animal collect ever increasing pieces of information about the rewarded behavior. After some trials, the animal will be very well informed about which behavior is appropriate. In the case of punishment, the outcome is the reverse. Since the animal will become less likely to perform the behavior that preceded the punishment, it will not be able to correct its possibly faulty view of the punishment. This implies, of course, that punishment is not a very good way to control the behavior of others.

5.12 Conclusion

This chapter presented a general reinforcement module that can be used as part of a number of learning systems. It is important to realize that although this learning module can be used generally, it is not a general learning mechanism in the sense that it solves all learning problems. It is only when this module is incorporated in other systems and connected to various behavior modules that it has any real power. This power comes from its ability to *coordinate* behaviors, and not *create* behaviors (Bolles 1984).

Although it can, in principle, learn arbitrary connections between stimuli and reponses, this is not the way the reinforcement module should be used. In chapter 8, we will see that learning of this type is highly inefficient and is only useful within the context of a larger system. It is also necessary to modify the reinforcement module, depending on where it is used in a system.

We have seen how similar principles can be used to explain a large number of learning paradigms. Many properties of classical and instrumental conditioning are very similar, which makes it possible to use similar models for both cases. The question remains as to whether the same mechanism is responsible for these two types of learning in real animals.

Learning is only one of the determinants of behavior, however. Another important determinant is the current need of the animal. In the next chapter, we will consider how these needs change over time and how they influence learning and behavior. We, thus, explore the area of motivation and emotion.

Chapter 6

Motivation and Emotion

6.1 Introduction

Some concepts are quite useless in their broadest meaning. Motivation is such a concept. There is, however, a more restricted sense of the word which is useful or even essential in any theory of cognition. This is the idea of motivation as a central state reflecting the combination of internal needs and external possibilities. In the motivational system, these two sources of information come together and they are used to form a decision on what to do. The representation of this decision constitutes the *motivational state* of the animal (See section 3.2), which tells the animal what it should do.

In this chapter, we will try to investigate the role of motivation as a determinant of behavior and learning. We will try to show that motivation and emotion plays an important role in cognitive processes which has often been overlooked in the past. The traditional dichotomy between cognition and emotion is probably responsible for the lack of interest in motivational theories within cognitive science (LeDoux 1995). This is unfortunate for a number of reasons.

First, there seems to be no clear distinction between cognition, motivation and emotion in our daily activities. For example, the planning of a vacation is usually considered to be a cognitive activity. Every step of the planning process is guided by one's ideas of a good holiday, one's hopes, fears, etc. Without this side of the planning process, the vacation would hardly be worth planning, since it could only result in any of an infinite number of equally valued plans, none of which would fit

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

one's requirements for a good holiday. It appears that all thinking is biased by emotive values in this way.

Next, cognition cannot be understood without reference to motivation. Higher cognitive processes like categorization and problem solving only play any role when they are used to fulfil some goal of the individual, that is, only when they are *motivated*. This is unfortunately an aspect of cognition that is often ingnored.

Last, it is important to realize that motivation precedes cognition. Cognition depends on the motivational system, whereas the motivational system can operate without cognition. This point will be developed further in section 6.8. We will also argue that the role of emotions is to tell the animal what it should have done, when its expectations are inaccurate or its behavior is unsuccesful.

We will see that the motivational state has two important properties: First, *the motivational state is central* to the whole organism. This means that at all times, the organism is influenced by one single motivational state which selects a single engagement. Second, *the motivational state is dynamic*. As the internal needs and external possibilities change or are reevaluated, the motivational state may change at any time.

6.2 The Determinants of Motivation

Learning is just one determinant of behavior. The other important determinant is motivation. Some early theories did not consider motivation as a source of behavioral direction. Instead, they focused on learning alone. For example, the experiments of Pavlov (1927) did essentially ignore the role of motivation. This is curious, since the classical association of a bell[1] with salivation could only be learned when the dog was hungry. A satiated dog simply does not show the salivation reflex.

In Thorndike's experiments, the role of motivation was even less clear. The cat in its box was given food when it had managed to escape, but this was more or less an afterthought. It was never clearly established whether the cat tried to escape because it wanted food or because confinement is aversive to the animal.

The role of motivation compared to learning was clearly shown in an experiment by Clark (1958). A number of rats were trained to press a bar in order to obtain food. The rats were then placed in different groups and deprived of food for between 1 and 23 hours. The longer the rats had been deprived of food, the faster they pressed the bar to obtain it. Since all the rats had received equal training, the differences could not be due to differences in learning. Consequently, the experiment shows that motivation, or in this case, the deprivation of food, plays a significant role in determining behavior.

---

[1] In many of Pavlov's original experiments, the sound of a metronome was used rather than a bell.

Hull (1943) proposed that deprivation induces an aversive state in the organism. This state was called a *drive*. In his conception, drive increases the overall level of arousal in an animal. Drive was thus considered as a property of need states which motivates behavior. Hull considered drive to have an overall effect on behavior in that it energizes it, but does not determine what behavior is performed. This is usually called a *generalized drive* and has now been replaced by the concept of arousal (Gray 1975). The theory can be summarized in Hull's law,

$$_SE_R = {_SH_R} \times D \times V \times K \tag{6.2.1}$$

where D is the drive level, V the stimulus intensity, K the incentive motivation — a concept we will return to below — and $_SH_R$ is the habit strength that results from learning. They all multiply to produce the *reaction potential*, $_SE_R$, which is a measure of the likelihood that an animal will produce a particular response, R, given a stimulus, S.

Later studies have shown that the role of drive is to *selectively* increase the frequency of behaviors likely to reduce the drive[2] (See Bolles 1967). In Hull's formulation an increased drive caused by deprivation of food would increase the likelihood of any behavior, not only of eating. Taking this into account, we must relate both the drive and the incentive to a particular type of motivation and Hull's law changes to,

$$_SE_R = {_SH_R} \times D_m \times V \times K_m. \tag{6.2.2}$$

Here, D_m is a specific drive, such as hunger, which relates to a particular motivation, m, for example motivation to engage in food seeking behavior or eating. The incentive motivation must also be related to the particular motivation that is indicated by the subscript m. Later studies have also shown that the interaction between deprivation and learning may be more complicated than multiplication (Bolles 1969), but the basic idea is very sensible and as we will see below, the concept of motivation can be used to great utility as the basis for a theory of choice.

Another possible extension, which we will not investigate further here is to use vectors for the representations of drives. This makes it possible to have a whole set of motivational states which represent, for example, hunger (McFarland and Sibly, 1972). This corresponds to the idea that one can have hunger for different types of foods.

6.3 Drives

Drive is probably one of the vaguest terms in psychology. It has been used in many different meanings, some of which have gained a very bad reputation. The use of drive that is intended here is as a state or value representing the urgency of a be-

---

[2] Note that this does not imply that it necessarily is the reduction of drive which controls the learning process.

havior. In some cases, a drive relates to a physiological state produced by a deprivation or increased levels of food, a hormone etc. In other cases they relate to the presence of a noxious stimuli such as a loud noise, pain, heat, etc. In these cases, drives are equivalent to need states, but we will also allow constant drives which compete with these more usual drives. Constant drives represent the relative importance of different engagements which do not change over time.

Drives, thus, have the role of representing the relative importance of an engagement at a certain time. This implies that a creature must have one, or possibly many, drives for each of its innate engagements. We can identify at least six different types of drives.

Homeostatic Drives The first type of drive that we will consider is generated by violation of homeostasis. This includes, for example, hunger and thirst, but also the responses to heat and cold. Violation of homeostasis induces a drive signal into the motivational system. The probability of an engagement increases with larger deviation from the optimal level but the drive signal does not directly activate a behavior.

It is instructive to compare this notion of drives with the behavioral model proposed by Powers (1973). Powers' model is built on the assumption that the present action of an organism is a function of its present perceptions and an internal reference perception. These are compared to generate an error signal which facilitates behavior. This is exactly the idea that the discrepancy between the present state of the organism (the present perception) and its desired state (the reference perception) drives behavior. The idea of an error signal is very similar to the idea of a drive. (Compare this idea with the discussion of approach behavior in chapter 4).

In Powers' models, this idea is generalized to hierarchically connected comparators. The comparator on each level generates the reference perception to the system below it. What starts out as a mechanism for maintaining homeostasis has turned into a behavioral control system. In the process, the motivational system has disappeared and so has the ability of the system to select among different engagements. Though elegant, Powers' model cannot explain the selection of one engagement among many when several discrepancies are present. However, it deserves recognition as an early and thorough system model of horizontal decomposition of competence (Compare Brooks 1986, and Schnepf 1991).

Powers' model treats behavior as a function of two variables: the present perception and the reference perception. We want to argue that three variables are needed: the present perception, the present motivational state and the 'reference perception'. However, the 'reference perception' needs not be explicitly represented since it is essentially fixed by evolution.

Discrepancies can thus be detected directly without reference to the optimal state of the organism. Thus, the comparators do not exist in a physical sense and the reference perceptions are not really perceptions at all. Yet, the organism behaves as if they were present.

Also, as we have already discussed, the role of the present perceptions is not to determine the violation of homeostasis, but to modulate the motivational state. This is especially important in, for example, sexual and maternal behavior. Motivation for these are not brought about by lack of homeostasis.

In the following, we will stick to the simplistic model of homeostatic drives described here, but there are admittedly a number of extensions which could be made. The most important extension is that most animals do not eat because they are hungry, but rather to avoid becoming hungry (Collier 1983). This means that an animal must be able to anticipate its future needs or alternatively defend their food intake rather than homeostasis. We can still use the notion of drive in these cases by allowing it to have a less direct relation to immediate physiological need.

Noxious Drives We may also consider the signals from noxious stimuli as drives. This includes, for example, the sensation of pain. Since noxious signals will also constitute external incentives, it is possible to consider the drive for avoidance of noxious stimuli as constant. As we will see below in the discussion of choice mechanisms, there is reason to include constant drives of this types, since it makes it easy to model individual dispositions of an animal as the levels of these constant drives. For example, an animal with a lower level on its 'pain avoidance' drive, will appear to be more resistant to pain than one with a higher level on this drive since it will engage in avoidance more seldom and with less vigor.

Cyclical Drives There are a number of cyclical drives which vary with the time of day or the year. These are not directly controlled by an internal or external stimuli. Instead they are generated by an oscillator. This oscillator in turn is influenced by external stimuli such as the length of the day, odors, or the amount of light in the environment. In this group of drives we find, for example, tiredness and wakefulness, the sexual drives of most animals, and migratory drives.

It is interesting to consider to what extent hunger drives are also of this type. Perhaps hunger is also controlled by an internal oscillator which locks to the cycle which decides the time when an animal usually gets hungry and eats. A mechanism of this type would greatly simplify the anticipation of future hunger (Collier 1983), but we will not consider it further here. It could also potentially account for fluctuation in eating behavior depending on the time of day or the season.

Default Drives The fourth type of drives can be called default drives since they only influence the animal when no external cue is present which commands the animal to do something else. The most important drive of this type is concerned with *grooming* (Bolles 1984). Such behavior is performed when the animal has nothing else to do. Typically, grooming behaviors consist of a fixed-action pattern (See section 2.3 and 2.4). Activities of this type are also used as *displacement activities*, that is, behaviors that are generated when an internal motivational conflict blocks the execution of a more appropriate behavior.

Exploratory Drive Another drive which is similar to default drives in that it tries to activate behavior when the animal has nothing else to do is the exploratory drive. The role of this drive is to let the animal try out behaviors either at random or directed by some exploratory mechanism to learn about its consequences in a certain environment. The exploratory drive can also interact with the perception of unknown stimuli to produce exploratory behavior which is directed toward a specific stimulus. In chapter 7 and 8, we will see how the exploratory drive interacts with learning and orienting behavior to produce efficient search of a novel environment.

Anticipatory Drives Finally, there is reason to believe that, in higher animals at least, there exists a sixth source of drive signals. This type of signals are internally generated, but do not relate to any present need of the animal. They can influence any other drive and their purpose is to simulate a drive, typically a homeostatic one, which is not present. This is necessary for the planning of future needs (Gulz 1991). This type of drive will be further considered in chapter 9 in relation to planning and anxiety.

The anticipatory drive is similar to the default drives and the exploratory drives in that it does not need sensory stimulation to become effective. It can be held at a constant level and activate anticipatory planning behavior when the animal has nothing more important to do (See Gulz 1991, and chapter 9).

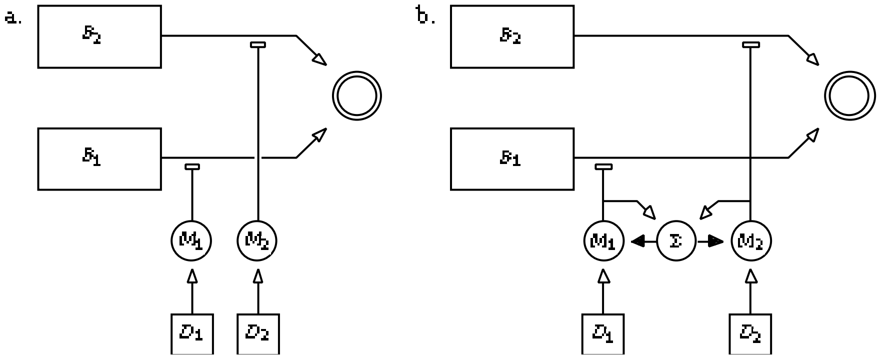


Figure 6.3.1 Motivation as a function of drive (D_1 and D_2). (a) The level of motivation (M_1 and M_2) interacts with the output from the behavior modules (B_1 and B_2) through multiplication. (b) Competition between the two motivational states allows only one behavior module to control behavior at a time.

Motivation as Competing Drives

We see that drives can be classified as *constant* or *varying*. In many cases, it may not be sensible to consider constant drives as drives at all, but in everyday language we talk about a curiosity drive much in the same way as a hunger drive. Although the mechanisms controlled by these two drives are very different, there is an underlying similarity in how the drives are involved and this makes it reasonable to use a similar formalization of both cases.

Let us now look at how the motivational state of an animal is changed by its drives. To a first approximation, the strength of a motivation, m, called M_m, depends only on a corresponding drive, D_m, which gives the following formula for the calculation of motivation:

$$M_m = D_m \tag{6.3.1}$$

In figure 6.3.1a, a simple network is drawn which can calculate motivation in this way. The motivational state influenced behavior by facilitating the behavior command sent from the two behavior modules (Werner 1994). In the right figure, competition between the two motivational states is introduced as described in section 1.4. As a consequence, only one motivation can be active at a time. This property of the motivational system is useful when the additive composition of the behaviors generated by the different behavior modules is not meaningful. When competition is included in the motivational system (figure 6.6.1b), the creature uses arbitration by central selection, which allows it to perform one behavior at a time (See section 2.3).

Figure 6.3.2 shows how the life space of an animal changes as a result of change in drive level for the network with and without competition. In the simulations, the approach tendency toward a goal of type m was multiplied with the corresponding motivation, M_m. As can be seen, the borders between the basins of attraction are moved away from the goal with the highest corresponding drive level as could be expected. The animal will consequently be more likely to choose that goal instead of the other. When competition between the two motivational states is added, only goals relating to the dominant motivation are perceived at all.

While this mechanism makes the creature attend to its most important need first, it does not generate optimal behavior. In chapter 7, we will investigate how an attentional mechanism can be added which lets the creature weigh together its needs with the psychological distance to the various goals. With an attentional mechanism, the creature can select the best overall goal without sacrificing its ability to perform one behavior at a time.

6.4 Incentive

Another of Hull’s contributions to this area includes the idea that motivation is determined by two factors. As described above, the first is the need state or drive.

The second is *incentive*, that is, the presence of an external stimulus which predicts a future reduction of the need of the animal. For example, the presentation of food would constitute an incentive to a hungry animal. By the process of secondary conditioning, other stimuli which predict food can also become incentives as described in section 5.6.

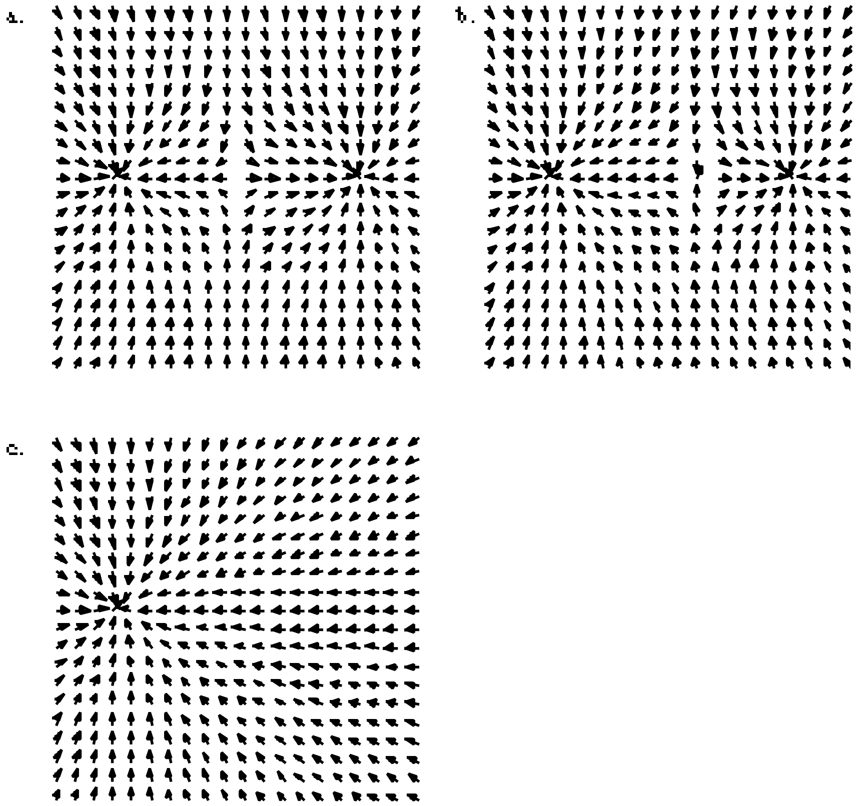


Figure 6.4.1 The life space of a creature with different motivational conditions. (a) The drives corresponding to each of the goals are equal, that is, $M_0 = M_1$ (b) The motivation for the left goal is larger than the motivation for the right one, that is, $M_0 > M_1$. This makes the left basin of attraction larger. (c) Competition is added which makes the goal with the least motivation disappear entirely.

Schachter (1970) describes a number of interesting studies concerning the distinction between internal and external cues as determinant of behavior. It appears that different individuals are more or less likely to base their actions on internal or external cues respectively. A number of experiments show a striking difference between the normal and obese subjects. While in normal weighted subjects, their self-reports of hunger feelings coincide closely with stomach contractions recorded by a gastric balloon, obese subjects show no such correlation. Instead, their eating habits seem to be almost exclusively controlled by external cues, or as formulated by Buck (1988), "[t]hey are, in effect, on a 'see-food' diet: if they see food, they eat it".

Although the theory is not without its problems, there exists a number of experiments which show that obese people, and rats for that matter, respond more readily to external stimuli than to internal needs. This is true, not only for food related cues, but for all types of external stimuli. If nothing else, these studies surely motivate the distinction between drive and incentive.

We may distinguish between two types of incentives. The first type can be called *external incentive*. This type of incentive is directly generated by an external object such as food. In section 1.3, we introduced yet another determinant of motivation which was called *internal incentive*. This third determinant should be distinguished from external incentive in that it is not directly derived from the external goal stimulus, but it must be generated by a more complex internal process whether this be the result of secondary conditioning, planning or even some innate structure.

	Primary	Secondary
External	Fish	Candy
Internal	Water	Candy Store

Figure 6.4.2 Examples of the four types of incentives for an animal which has innate mechanisms for the recognition of fish and water and the relation between them, but must learn the valence of candy and where to get it.

External incentives are usually closely connected with approach and consummatory behavior since the goal object is present when an external incentive is generated. Secondary incentives, on the other hand, are related to instrumental appetence behavior since the goal object is not yet present.

It is also possible to distinguish between *primary* and *secondary* incentives. Primary incentives are available to the animal before any learning takes place, while secondary incentives are the result of learning. Typically, primary incentives are external and secondary incentives are internal. We may thus distinguish between the four types of incentives exemplified in figure 6.4.2

Taking incentives into consideration, we can now extend our motivational system further. The strength of a motivation can now be calculated by the formula,

$$M_m = D_m(I_m + E_m) \qquad (6.4.1)$$

where, the three inputs are the drive, D_m, the internal incentive, I_m, and the external incentive, E_m. These three values are all functions of the internal and external stimulus situation of the animal. To allow constant drives, we must also include a constant, d_m, which is added to the function D_m. We also add a constant to the internal and external incentives called c_m, which allows the level of motivation to be non-zero even when both incentives are inactive. The calculation of the strength of a motivation is now given by,

$$M = (D_m + d_m)(I_m + E_m + c_m) \tag{6.4.2}$$

The corresponding neural network is shown in figure 6.4.2.

This view of motivation is very similar to the idea of inter-behavior bidding discussed by Sahota (1994). All behaviors, or in this case, engagement modules, can be considered as bidding for the control of the creature at any time. The engagement which places the largest bid according to equation (6.4.2) will be allowed to control the creature. As pointed out by Minsky (1987), a bidding scheme of this type can easily give rise to oscillations in the choice of engagement if drive alone is the only factor influencing the choice. In the next section we will see how such oscillations can be prevented.

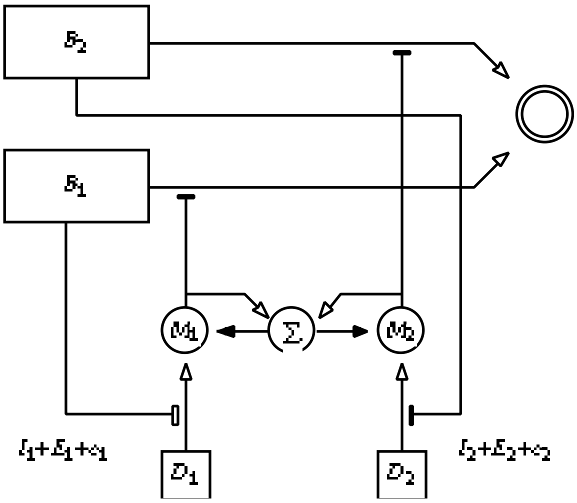


Figure 6.4.3 A motivational system which receives incentive signals from the two behavior modules. These signals are composed of the three factors, internal incentive, *I*, external incentive, *E*, and a constant factor, *c*, and gate the signals from the drive receptors.

6.5 Choice and Persistence

There is one large problem with the architecture shown in figure 6.4.2 which is not immediately obvious. Since the creature will always choose the behavior with the largest level of motivation, oscillations can easily occur. Suppose, for instance, that the animal is both hungry and thirsty, but slightly more thirsty than hungry.

Assume further that food and water are readily available, but at different locations which require that the animal moves from one place to another to eat or to drink.

Since the creature is more thirsty than hungry, its first behavior will be to approach the location of the water. It will then drink a little and its thirst drive will decrease slightly. This will cause the hunger motivation to win the competition and the creature will set of for the food. Once it has eaten a bite, the thirst will be stronger than the hunger, and it will set off for the water again, and so on. This is obviously not a very clever strategy since most of the time will be spent on the way between the food and the water.

There are two solutions to this problem. The first is the incentive mechanism discussed above (See also McFarland and Bösser 1993). This will make the creature more hungry the closer it comes to the food, which will make it stay at the food longer than without the incentive signal. The incentive mechanism, thus, sets up a positive feedback-loop. This is not enough, however, and in many cases an additional mechanism is needed.

In this solution, two separate opposing systems are introduced for each drive. For example, to hunger, we add a system for satiety. When the animal becomes hungry, the hunger motivation is held high until it becomes inhibited by the satiety system. This will let the creature eat until it get satiated and not only until it is no longer hungry. By including this type of hysteresis, the eating behavior will be more persistent.

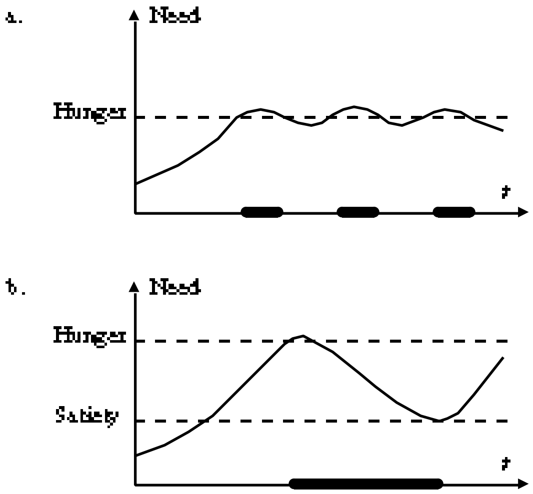


Figure 6.5.1 The effect of having two opposing systems for the hunger motivation. (a) Only one system is used which increases the level of hunger when the need raises above a set point. This will result in oscillations which cause short bursts of eating behavior. (b) The effect of using two opposing systems instead. A hunger system activates the eating behavior and a satiety system inhibits it again when the need has decreased sufficiently.

In figure 6.5.2, a simple neural circuit is shown which can generate a persistent hunger motivation as the one shown in the bottom graph of figure 6.5.1. A single sensor is used for both hunger and satiety. In general, different sensory systems could be used for the two opposing systems. This is shown in figure 6.5.3. Note that in both networks, the original drive signal is added to the persistent drive level to let the choice in the motivational system reflect the actual level of the drive to some degree. In the hunger system of real animals, many systems of this kind interact with each other to produce the feelings of hunger or satiety. It has been suggested that centers in the hypothalamus play a role similar to these systems (Teitelbaum 1961) although this view is now recognized as too simplistic (Grossman 1979, Collier 1983).

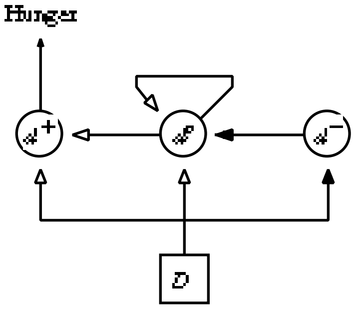


Figure 6.5.2 A network which generates persistent hunger until satiety occurs. The middle node, d^p, becomes active when the drive level raises above a certain hunger level and represents the persistent component of the drive. This node keeps itself active using a recurrent connection until it is shut off by the satiety system, d^-. The satiety node is inhibited until the drive level decreases below the satiety level. The persistent hunger is added with the original drive signal at d^+ to produce a hunger signal which is persistent but still reflects the absolute drive level.

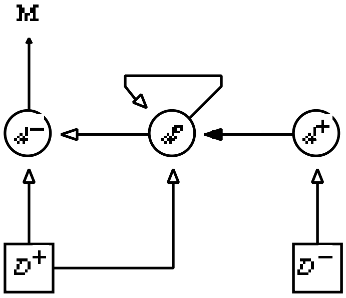


Figure 6.5.3 The general architecture of a network for calculation of persistent motivations. The drive sensor D^+ increases motivation and becomes persistent when it increases sufficiently as in figure 6.5.2. The sensor D^- activates an inhibitory system which shuts off the persistent motivation.

6.6 Optimal Choice

How should a motivational system be constructed to let the creature make the optimal choice of behavior with respect to psychological distances and expected rewards? The answer to this question depends on what the overall goal of the creature is. We will consider two possible answers here. The first is that the creature should try to maximize the reward it receives and the second is that it should try to minimize the risk of starvation and death.

If we take the first position, the motivational system needs to calculate the expected reward for each potential action which can be performed in a certain situation. In chapter 5, we saw how instrumental conditioning could be used to estimate the potential reward in each situation with respect to a specific goal. When many goals are present, this mechanism will instead calculate the potential reward with respect to the closest goal of a certain type.

For example, one reinforcement module may deal with one specific hunger, while another may deal with another hunger or, for example, thirst. The choice to be made in the motivational system concerns what type of behavior the creature should engage in. For example, should it activate its engagement module for eating or drinking? Once the choice has been made, the individual engagement systems take care of the actual execution of the behavior. This suggest that internal and external incentive should be calculated by the engagement modules and then sent to the motivational system where they should influence the choice of behavior.

Let us first assume that the drive levels for two engagements, d_A and d_B, are equal. This implies that the creature should select the engagement which will reach its corresponding goal with the least effort. That is, the creature should select the engagement for which the discounted reward is highest. Recall from section 4.2 and 5.8 that the value of a reward, R_A, received at a goal, g_A, is computed by multiplying it with the discounted cost, $c(z,g_A)$, of moving from the present location, z, to the goal, g_A. If g_A and g_B are the goals corresponding to the two motivations, the creature should select engagement A if,

$$\delta^{c(z,\,g_A)}R_A > \delta^{c(z,\,g_B)}R_B. \tag{6.6.1}$$

and engagement B otherwise.

This implies that when learning is included in a creature, incentive is equivalent to primary or secondary reward. Primary incentive can be generated from the primary reward and secondary incentive from the secondary reward. Note however, that incentive may be present even without learning which means that incentive and reward are not identical concepts. Given this similarity, we can define goal situations as *primary motivators* and stimuli which generate secondary incentive as *secondary motivators*.

How can we extend this choice to situations where the drive levels are not equal? We could se above, in equation 4.4.2, that the incentive signals should be multiplied with the drive signals. The choice situation above will then be transformed into the following. Select engagement A if,

$$d_A \delta^{c(z, g_A)} R_A > d_B \delta^{c(z, g_B)} R_B, \qquad (6.6.2)$$

and B otherwise. We see that the motivational system described above calculates the optimal choice in this way.

There is one question that we have not considered yet, however. How should the drive level reflect the needs of the creature to make its choices optimal? This question cannot be answered generally but must depend on how the different engagements of the creatures relate to each other. One principle that can be formulated, however, is that an animal should minimize its loss rather than maximizing its gain. These two strategies may appear identical, but there are many cases where they are different. This is especially true when the outcome of a behavior is uncertain. To survive, an animal should avoid making choices which could result in starvation even if the risk is very small. No matter how much food it can receive by performing a potentially lethal behavior, this gain should never outweigh the risk, if some other possibility exists which will keep the animal alive with almost absolute certainty.

On the other hand, an animal should always choose to eat if possible when it is on the edge of starvation. This implies that drive should not necessarily be a linear function of need. Figure 6.6.1 shows how a hypothetical hunger drive may increase drastically when the need of the animal passes above a certain point.

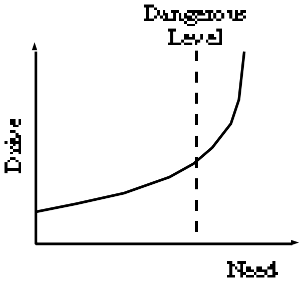


Figure 6.6.1 Drive as a function of need. The drive signal increases drastically when the need reaches dangerous levels to make certain the animal chooses the appropriate engagement.

Principles like this may be the mechanism between the apparently irrational choices made by human subjects in many experiments. Subjects will prefer to play a game where the gain is smaller, but certain, rather than a game where the gain is higher but at the risk of not winning anything at all (Pitz and Sachs 1984). If the price was food and the subject was a hungry animal on the edge of starvation, the choice would in fact be rational. However, since the modelling of these types of choice mechanisms require a representation of certainty, we will not consider them further here.

6.7 Emotion

The description of motivation above raises the question of where emotions fit in? A definition of emotions which fits nicely with the view of motivation presented here is that emotions are “states elicited by reinforcing stimuli” (Rolls 1986, Gray 1982). This definition results in a view of emotions where they can be categorized along four dimensions which are arranged according to figure 6.7.1.

We have already encountered these dimensions in chapter 5. We saw that they result from the comparison of actual rewards or punishment with expected ones. This implies that expectations are necessary for emotions along the rage–relief dimension. Without expecting a rewarding situation, how could one get disappointed? The same is true about relief. If we did not expect a punishing event to take place, we would not feel relief when it is omitted.

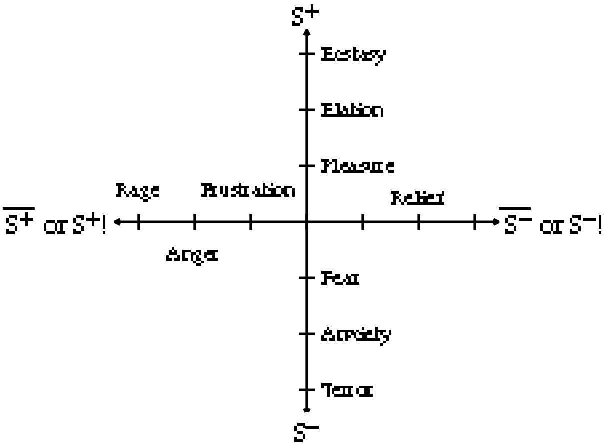


Figure 6.7.1 The dimensions of emotion. The presentation, omission or termination of reinforcing stimuli form the basis for a number of emotional states (Adapted from Rolls 1986. See also section 5.2 and figure 5.2.1).

As pointed out by Rolls (1986), the type of emotion is also dependent on what behaviors can be performed at a given time. When active behavior is possible when a rewarding stimulus is omitted, it may result in anger. If only passive behavior is possible, sadness or depression, may result. This aspect of emotions can be compared with the role of the applicability predicate in selecting between preparatory and consummatory behavior (See section 5.3).

For the pleasure–fear dimension, two mechanisms are possible. In the first case, the emotional level directly depends on the intensity of the reinforcing stimuli. In the other case, the emotion also depends on expectations. In this case, the emotion is generated by the reinforcement which results when the reinforcing stimulus is presented. No reinforcement is generated when the reward or punishment is already expected. This implies that emotional states will be weaker the more expected the reward or punishment is. According to this view, a reward which is entirely expected will not result in any pleasure at all.

When emotions are identified with reinforcing signals rather than stimulus intensity, an animal needs to increase the reward it receives to evoke an emotional state of constant intensity. Constant ecstasy will, thus, require a constantly increased level of the rewarding stimulus, a fact well known to drug addicts where this property takes hazardous proportions.

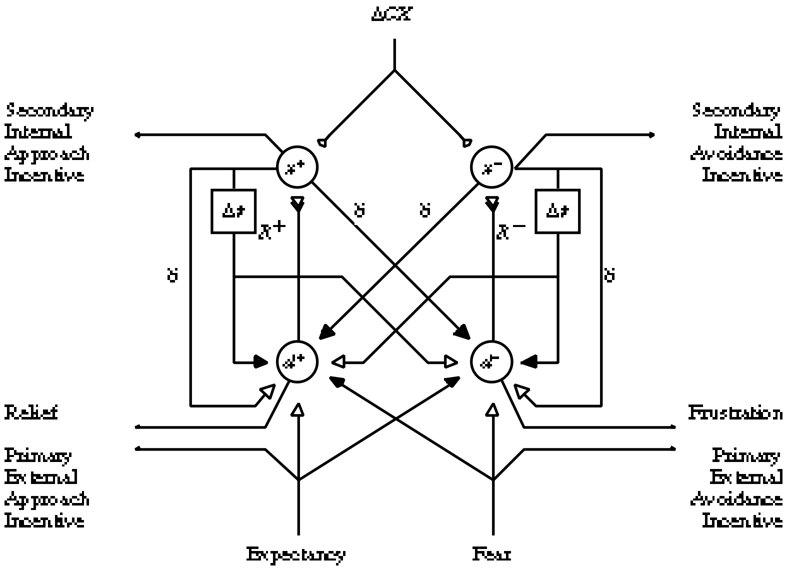


Figure 6.7.2 Emotion as reinforcement. The different states of the reinforcement module can be interpreted as emotions. The primary reinforcers induce fear and pleasure (or expectancy). The omission or termination of fear will induce a state of relief. Omission or termination of an appetitive state induces frustration.

In less dramatic contexts, a creature that strives for pleasurable experiences instead of a low level of motivation will constantly need more reinforcing stimulation. Since this is not always desirable, the creature needs some mechanism which will prevent events from becoming reinforcing when they are dangerous. For example, food should not be reinforcing to a satiated animal. Indeed, too much eating does result in pain. For other motivations, such as exploration, no such system is necessary.

An important property of emotions is that they are motivating in two different senses. In the first sense, they are motivating in the future, since they increase the likelihood of a certain behavior at a later time. For example, a creature, which has been rewarded when it has performed a certain behavior, will generate a larger incentive signal the next time it finds itself in the same context. This in turn will motivate the behavior.

However, emotions also appear to be motivating directly, since a reinforcing stimulus often causes a specific behavior to be performed. For instance, the presentation of food will motivate an eating behavior. This consequence needs not to be generated from the reinforcement signal which constitutes the emotion, however. It may merely coincide with it. It is quite possible for this effect of the reinforcing stimulus to be present although no emotional state is generated.

Figure 6.7.2 shows how the different emotions relate to the reinforcement module introduced in chapter 5. This module receives three types of inputs and calculates six types of emotional signals. The two primary inputs comes from perceptual systems which identify fear and expectancy respectively (Compare Panksepp 1986). Secondary input is generated by the contextual input which tells the reinforcement module about prior rewards and punishments. There are three outputs on each side of the reinforcement module. One for primary extrenal incentive, one for secondary internal incentive, and one for the omission or termination of the reinforcing stimulus for the opposing system.

When a punishing situation occurs unexpectedly, it will generate incentive signals for avoidance behavior and start the learning process in the reinforcement module. This state is called ‘fear’ (Rolls 1990). Stimuli that predict the punishing stimulus will generate secondary avoidance incentive, that is, passive avoidance or behavioral inhibition. When such a stimulus is presented, the resulting state is called ‘anxiety’ (Gray 1982). When the expected punishment is omitted, the resulting state is called ‘relief’.

Figure 6.7.3 shows how the signals from the reinforcement module influence the motivational system. The various outputs are used as incentives or drives for different motivations. On the appetitive side, internal and external incentive is multiplied with the drive signal that indicates the current need of the creature. The product of these two signals is used to activate motivation m_1. When this motivation wins the competition, the creature will approach and consume the reinforcing stimulus. On the aversive side, the primary incentive facilitates motivation m_3, which is assumed to have a subthreshold resting activity which corresponds to a constant drive. If the incentive is high enough, the creature will activate an active avoidance behavior.

There are also two motivational states corresponding to the two emotional states of relief and frustration. Both these emotions act as drives for particular motivations. Motivation m_1 which is associated with relief makes the creature act in whatever way is appropriate when it feels relieved (perhaps it smiles). When a stimulus appears that predicts punishment, motivation m_4 is activated which generates passive avoidance and when the creature becomes frustrated, motivation m_5 is activated which may cause the creature to attack the cause of the frustration, or in other cases, an innocent bystander. Since there is no equivalent of behavioral inhibition on the expectancy side of the reinforcement module, the effects of fear and expectancy are not symmetrical.

It is very common for the immediate motivation that accompanies emotions to generate some form of communicative signals, but since we will not deal with social interaction in this book, we will not consider them further here.

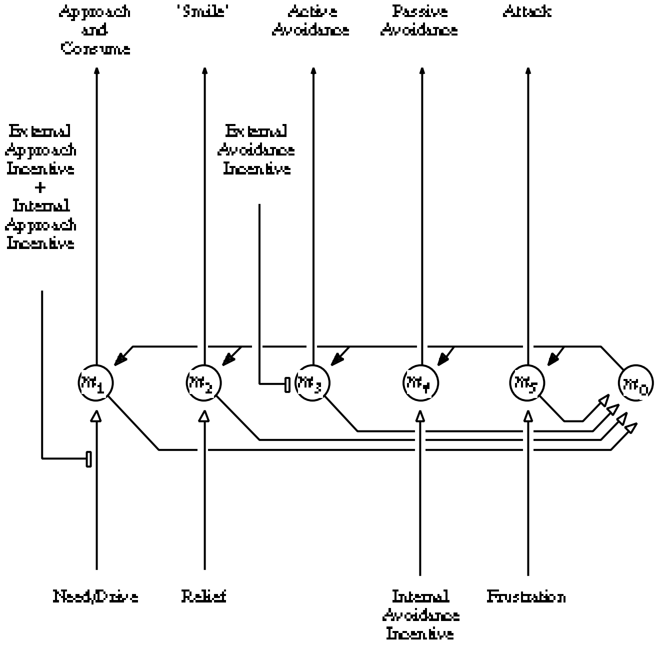


Figure 6.7.3 Motivational states influenced by the reinforcement module in figure 6.7.2. Presentation, termination and omission of stimuli activate motivational states for agression and expressions of emotion. These motivations compete with the more basic ones for appetitive and aversive behaviors.

The property of motivational states to cause frustration or relief when it is changed has been called the motivational *rebound effect*. This effect rests on the fact that many motivational states seem to come in pairs: tiredness and wakefulness, hunger and satiety etc. When the need corresponding to the motivational state is satisfied, the opposite motivational state is activated. The phenomena have been much studied, for example, in relation to drug abuse (Solomon 1980). The introduction of the drug is followed by a state of happiness, but when the drug is removed the opposite feeling arises. The same is true of fear and relief. After being exposed to a situation of extreme fear, the termination of fear does not result in a neutral state but rather in euphoria.

There is, however, more to motivation than can be included in the framework presented above. In agreement with self-attribution theory (Schachter 1964), what a creature may *experience* as emotions is the result of its internal categorization of the motivational system and its effect on behavior. The motivational system, as such, does not produce emotional experiences. The only function of the motivational system is to direct behavior now and in the future. It may very well produce the external attributes of emotions such as a smile or tears, but the experience of emotion is the result of an introspective process. This places the experience of emotion clearly within the area of cognition.

This view also implies that emotions must be learned from experience. How hunger effects behavior in the direction of food is not initially known to the animal. It is learned from its own behavior and perhaps later learned to be ignored or suppressed. If emotions are viewed in this way, it is not too remote to place love and hate side by side to hunger and tiredness. They must all be experienced and learned before they can become emotions. They differ in complexity only but not in nature.

To conclude it is possible to consider motivations as states which tell the organism what it should do at a certain time based on its internal needs and external possibilities. Emotion, on the other hand, is concerned with what the animal *should have done*. When the reward or stimulus situation of the animal is unexpected, an emotional state is activated. This state has two functions. The first is to control learning which lets the animal cope with the preceding situation in a better way the next time it occurs. The second function is to motivate specific emotional behaviors.

6.8 The Roots of Motivation

In this section, we will present a number of evolutionary steps that may have lead to the type of motivational system described above (See also Balkenius 1993). To do this, we have to start with some design principles that seem reasonable from an evolutionary point of view. The first pair of principles is reduplication and variation. Reduplication is a process which makes copies of an already existing structure and variation is the process whereby existing structures can be changed. We want to propose that it is possible to find three architectural principles which superimpose on each other during evolution. Based on their underlying architectural principle, we can define three classes of systems: (1) the subsumption architecture, (2) the centralized control architecture, and (3) the layered architecture.

These architectural principles do not directly describe the motivational system, but the organism as a whole. Each class is necessary for the next to evolve and each subsequent class of systems have a competetive advantage compared to systems in the previous class. This ensures the adaptive value of architectural changes which moves a species towards a higher class. The rest of this section describes the different classes in detail and identifies the necessary steps from one class to the next.

The Subsumption Architecture

The simplest system will possess a *subsumption architecture* (Brooks 1986). Such an architecture consists of a number of distributed systems which control different behaviors. They can, but need not interact, with each other. A subsumption architecture typically consists of a number of layers which can subsume each other. For example, a lower layer can produce walking behavior while a layer on top tells it to walk forward, turn to the left or to the right. Within AI, this approach to the construction of intelligent robots have gained an increasing popularity during the last few years (Beer 1990).

An architecture of this kind has three main features. First, it is non-representational. There do not exist any representations of the external world. Instead, the world is used as its own best model (Brooks 1991a). Second, the creature is reactive. It reacts directly to stimuli in the world. For example, a creature can avoid an obstacle simply by changing the motor patterns of its legs when its whisker contacts with the obstacle. Third, behavior is distributed. Different behavioral modules operate independently of each other to a large extent.

It is now possible to give examples of some simple creatures which use a subsumption architecture. We will try to keep the presentation as simple as possible to emphasize on the essential features of each architecture. Of course, real creatures are much more complex.

Our simplest system will simply consist of sensors and effectors. The sensors are connected directly to the effectors and either activate or inhibit them. Typical examples of such creatures are the vehicles invented by Braitenberg (1984). The simple control system of the vehicle is sufficient to produce taxic behavior. In *protozoa*, the entire control system can reside in a single cell. In larger animals, however, a simple nervous system is required to send the signal from the sensors to the effectors.

In a larger creature, the need arises for a connecting inter-neuron between the sensor and effector. This step does not change the behavioral abilities of our creature but is a requirement for its greater physical size. It is also essential to the future development of the species, as it is the origin of the nervous system. The next two classes can emerge in any order during evolution, but the two steps are both necessary for the evolution of the motivational system.

The introduction of an inter-neuron makes it possible for the sensory signals to interact before they reach the effectors. For example, two sensory signals can converge on a single inter-neuron which will then detect the conjunction of two sensory events. It is also possible for one inter-neuron to inhibit or activate another inter-neuron to produce complex behavior. In these systems, the inter-neuron integrates information from several sources and transforms it in some way instead of just propagating it from one sensor to one effector. Many of the systems described in section 3.4 are of this kind.

A parallel development is the introduction of sensors which can be of two kinds: external and internal. This is the result of a bodily change that places some sensors inside the organism. An external sensor reacts on some external event or state while an internal sensor reacts on the internal state of the organism. For example, the internal sensor of the feeding system could react on hunger and the external system could react on food being present. If both sensors where signalling, an eating behavior would be triggered. Computer simulations have shown that such nodes can spontaneously emerge during evolution (Cecconi and Parisi 1993). This is the origin of drives and incentives as defined above.

As in the simpler systems, a number of controlling systems can, in principle, exist in parallel. Each system is governed by its own sensory signals which control its behavior regardless of the other systems. The robots presented by Brooks (1991) are essentially of this complexity. They can produce very complex behavior without any central control mechanism. The central argument of this section is that such an architecture cannot, however, be extended to include cognitive processes without including a central control mechanism.

The Centralized Control Architecture

The creatures described above were characterized by the lack of central control. This is changed in the next class of creatures which exhibit the simplest possible architecture capable of centralized decisions. This is a necessary step toward a higher cognitive system and the origin of the motivational system. A similar position is held by Sjölander (1992) who discusses the need for centralized representations of objects.

When the organism develop more complex behaviors, they will inevitably disturb each other. When this happens, it is essential that the different systems can inhibit each other in order to produce *one behavior at a time*. In a neural context, this can be achieved by *lateral inhibition* (See section 3.3). There exist three different forms of lateral inhibition which could produce the desired result: feed-forward inhibition, feedback inhibition and recurrent inhibition (Grossberg 1973). Feed-forward inhibition is primarily used in other contexts to compensate for different overall activity levels (for example in the visual system in the LGN). Feedback inhibition have been used in at least one model of animal behavior which was based on the theories of Lorenz and Tinbergen (See Schnepf 1991).

In the following, we will assume that our creature is using recurrent inhibition to select one behavior instead of the other as described above. The properties of recurrent inhibition are essentially the same as for feedback inhibition, but they are simpler to analyze. In a recurrent network where the lateral inhibitory connections are of the same strength, the behavior that receives the most activation will be selected. One layer collects the activation of each behavioral system and the next layer is responsible for the selection of the behavior with the strongest activation.

The Layered Architecture

The next evolutionary step is to include several levels of control. This is not to say that a layered architecture is not possible on the lower levels. Indeed, the subsumption architecture is typically constructed in a number of layers. However, the nature of those layers are very different from the layers discussed here.

At this final stage, the system can develop in a new way. While the basic subsumption architecture, together with the lateral inhibition, keeps the creature alive, evolution can experiment on the next layer of our architecture. One of the most powerful mechanisms invented by evolution is that of learning. A learning ability can evolve on top of the previous system once motivation has been included, since learning has an obvious relation to motivation as we have seen above. Learning needs the motivational state to detemine when a certain behavior is good. Without it, the creature could only learn that a behavior is good or bad, but could not relate it any of its needs.

The learning architecture makes it useful for the animal to develop a more advanced perceptual system. It can supply the learning system with more advanced perceptual cues. Mechanisms for categorization and classification will now give the animal an adaptive advantage. From these capabilities evolves the mechanisms, which are usually considered to constitute cognition.

6.9 Conclusion

Words such as drive, incentive, motivation and emotion have had many different meanings within different theories. Even today there is no consensus as to what, for example, an emotion is (LeDoux 1995). What is called drive in one theory may be called emotion or motivation in another. We have tried to use these word in a way that is, at least, consistent with each other while retaining as much as possible of their ordinary meaning. Since our main concern is artificial creatures rather than real animals, one cannot help feeling that these concepts are more complicated in real animals. Clearly, we are nowhere close to the complexity of human emotions.

We believe, however, that in taking a design perspective on motivation and emotion, we have shown that these concepts can be used to great utility in the design of artificial creatures.

We have seen that a creature needs a central motivational state which is responsible for the selection of one motivation at a time. This requirement is closely connected to the idea of separate modules for different engagements which is essentially a strategy for task decomposition (Tenenberg, Karlsson and Whitehead 1993).

The choice of motivational state is influenced by three factors called *internal* and *external incentive* and *drive*. The incentives tell the creature about the current possibilities of fulfilling a need, while the drive signal informs it about the urgency of that need. Incentives can also be classified as *primary*, that is, innate, or as *secondary*, that is, acquired. Primary incentives were mapped onto the concept of primary motivators, and secondary incentives were considered as secondary motivators. This implies a close coupling between perception and motivation as will be further developed in the next chapter.

The view of drives, which we have presented, assumes that there is one or several drives for each engagement of an animal. This is thus an attempt to resurrect one of the original meanings of the word drive where it was more or less analogous to the concept of an instinct. While it has been argued that the explanatory value of this drive concept is nil (See Bolles 1967), we hope to have shown that it is quite useful from a design perspective. This usefulness comes from the separation of the drive concept from the engagement it controls.

The idea of pure emotions relating to a small set of biological processes is very similar to the idea of engagement systems described above (See Plutchik 1991). The problem for any such theory of emotion (or motivation) is, of course, to identify exactly which these basic processes are (Ortony, Clore and Collins 1988). We have made no attempt to produce a comprehensive list of engagements, since we do not believe that any such general list exists. The set of engagement systems required by an artificial creature depends critically on what it is supposed to do, and so will its emotions.

It was shown how emotions can be seen as states caused by reinforcing stimuli. Depending on the nature of the stimulus, emotions could be categorized into four basic dimensions: pleasure, fear, frustration and relief. It was also shown how incentives could be generated by the reinforcement module.

As mentioned at the end of chapter 5, the perfect symmetry between the different emotions is a simplification. Since many different engagement systems are involved, it would be more accurate to depict emotions in a three dimensional space where the third dimension would represent the engagement. Some engagements may not include both the positive and negative side of the diagram shown in section 6.7. For example, an engagement system for aggression may not include the pleasurable side. Such an engagement system may, however, interact with other systems (Compare Konorski 1967).

It has been suggested that the amygdala controls the emotional processes which form associations with primary reward and punishment, while the orbitofrontal cortex may be involved in the detection of mismatch between actual and expected reward or punishment which results in frustration or relief (Rolls 1986, 1995). The amygdala receives inputs from a large variety of sources in the overlying temporal lobe cortex and other areas. Some outputs from the amygdala, as well as from orbitofrontal cortex, are directed towards the hypohalamus which is known to mediate motivational and emotional responses (Rolls 1986).

To summarize, the basic function of motivations is to tell the creature what it should do, while the role of emotions is to tell the creature what it should have done.

Chapter 7

Perception

7.1 Introduction

Although this chapter is called perception, we will make no attempt to cover this complex but interesting area to any great extent. Even for the simple modalities of our artificial creatures, this would probably require more space than this entire book offers. What we will do, instead, is to discuss what role perception plays in an organism and how it relates to the other areas we have already discussed. What is the role of perception in learning and behavior? How does it influence motivation and emotion? How is perception controlled by the motivational state?

Depending on our perspective, the mechanisms we will present can be considered to control either perception, associative learning or motor control. We will see that it is not possible to make any clear distinction between these processes. We will identify at least three different roles that perception plays in an organism[1].

The first is to characterize the situation in which the organism finds itself. Is it new or familiar? Does the animal know what to do in this situation? The characterization of the situation is used as the context in which behavior is performed. This is the origin of the contextual representation used in chapter 5. We will present a mechanism that creates new situation categories when they are needed. Such a

---

[1] Compare these functions of perception to the different types of stimuli that were used to influence associative learning in section 5.2 (See also figure 5.2.3)

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

mechanism is also necessary for stimulus–response learning when the stimuli involved are not innately recognized.

The second function of perception is to guide the animal toward or away from a place, situation or object. Which way should the animal move to reach an object not visible from its current location? We have already discussed this mechanism in stimulus–approach behavior. In this chapter, we will extend it to *place–approach* or *piloting* behavior as defined in chapter 2. The important mechanism here will be the ability to learn and perceive *configurations* of stimuli.

The third function of perception is to tell the motivational system about the external possibilities present at a certain time. This mechanism was introduced in chapter 6 in relation to motivation. The signals sent to the motivational system from the perceptual apparatus were called *external incentives*. They make it possible for the motivational system to select the engagement which has the greatest chance of success at a given moment.

A related phenomenon is the motivational influence on perception. A hungry animal will be more likely to perceive food related stimuli than a satiated one. This means that the motivational state must somehow be able to bias the perceptual process. We, thus, need to transfer information in both directions between the perceptual and the motivational system. We will argue that the interaction between motivation and perception is the basis for one form of attention.

7.2 Situational Categories

In chapter 4, we assumed that there were sensors for every stimulus relevant to the animal. This was, of course, a simplification. In an environment with some complexity, it would constitute a large burden for a creature to have innate recognition mechanisms for all potential stimuli, most of which it would never encounter. A much more economical solution would be to create new categories when unknown stimuli are enountered. This leaves us with a new problem. How do we know when a stimulus is unknown?

There are two distinct ways in which such a situation can be recognized. The first mechanism is based on the similarity between the present stimulus and the already recognized stimulus categories. If a stimulus does not match any of the known categories sufficiently well, a new category must be created.

The other possible basis for category formation is to look at the predictions made from a recognized stimulus. If these predictions match expectations, everything is fine. If not, a new category may have to be created. We will see that this second mechanism depends on an ability to form expectations by classical conditioning.

Categorical Learning Based on Stimulus–Prototype Similarity.

The model of categorical learning we will use is based on competitive learning. This learning method is based on a set of nodes that compete with each other for activation, as described in section 3.3. All nodes receive the same input pattern, and the rate of learning is proportional to the output of the winning node (figure 7.2.1). This output is connected back to the node and is used as a learning signal.

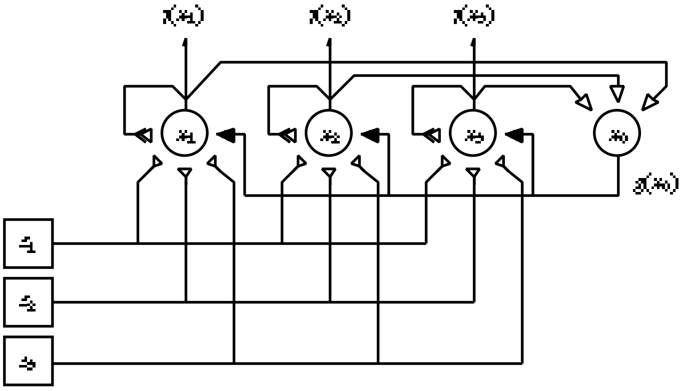


Figure 7.2.1 A competitive categorization network. Three category nodes $x_1 \ldots x_3$ compete for activation and will sort the input patterns at $s_1 \ldots s_3$ into three categories.

It is well known that a network of this type will establish the same number of categories as the number of competitive nodes (Grossberg 1976, Rumelhart and Zipser 1986). The set of weights on the connection to each competitive node acts as a *prototype* for its category. When an input pattern is presented, it is compared to the prototype of each category, and the output of each node is set according to the similarity between its category and the input pattern. The output of a category node will become larger with an increased similarity between its prototype and the input pattern. The competition between the nodes will then inhibit all nodes except the one with the largest activity. This winning node will, thus, signal the category of the input pattern.

While competitive learning has many useful properties, it also has a number of problems that must be addressed before it can be used for categorical learning in an efficient way (Grossberg 1976). The most important problem is that the learning is not incremental. The usual competitive network requires that all categories are trained at the same time. This implies that the network must relearn all previous stimuli every time a new stimulus is introduced. This is obviously not very practical for a creature which interacts with only one part of its environment at a time.

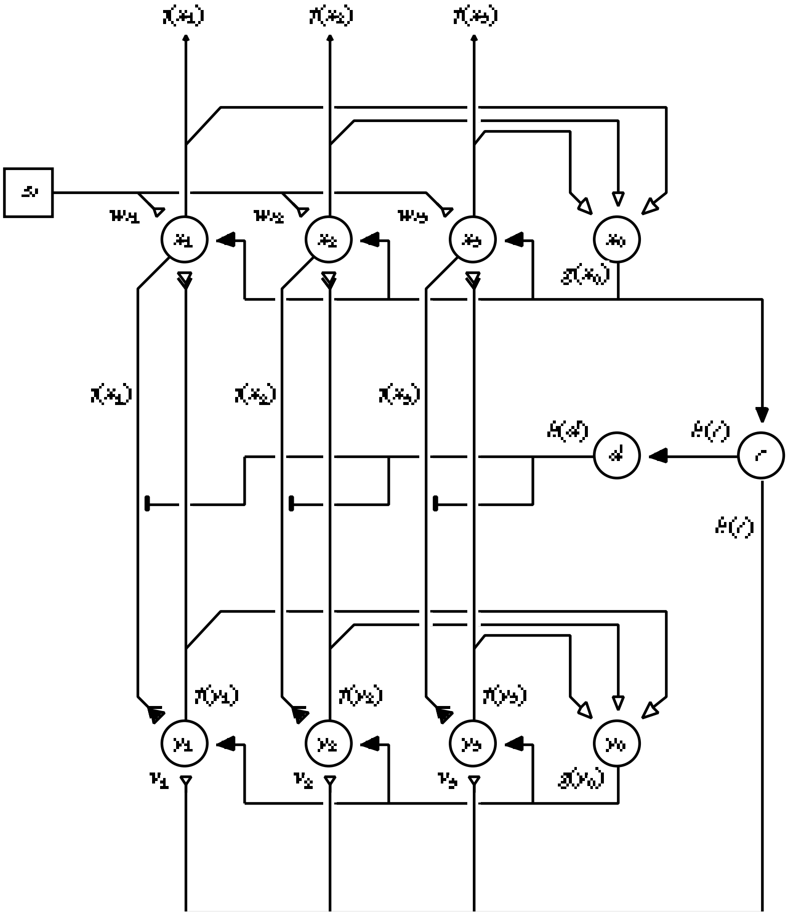


Figure 7.2.2 A competitive network with incremental learning. The network consists of two competitive networks, one for categorization at the top, and one for recruitment of new categories, at the bottom. The output from x_0 describes the similarity between the prototype of the best–matching category and the input. If this level is too low, the recruitment system will become active and create a new category in the uppermost competitive network. When a new category has been formed, the node *d* will be disinhibited and will consequently disengage the recruitment system. This will consolidate the memory of the new category. See the text for further explanation.

The ART[2] network described by Carpenter and Grossberg (1986) solves this problem by adding some extra nodes to the network that decides when a new category should be created. We will make a similar extension here, but will partly make

---

[2] The function of the network developed here will be most similar to the revised ART network for real-valued input vectors called ART II (Carpenter and Grossberg 1987).

some other design choices. The basic idea behind the ART network is that it is the difference between the sensory pattern and the best matching category that decides what type of learning will take place.

If the best matching category is sufficiently close to the input pattern, this category will change slightly in the direction of the current input pattern. Over time, this will make the category node respond best to the average of all input patterns the category has recognized. This average input will act as the *prototype* for the category. If the best–matching category is less similar to the prototype of that category, a new category will be formed. The segregation into these two types of learning guarantees that already established categories cannot be destroyed by later inputs and also allows one–shot learning of new categories.

We will develop a network that uses the same basic idea but operates in a different way (see figure 7.2.2). A recruitment network is added to the competitive network above, which handles the creation of new categories. Such a modification of the network for competitive learning was suggested by Rumelhart and Zipser (1986) and has also been implemented by Rolls (1990). It operates in the following way. The output from the node x_0 will give an indication of the similarity between the best category and the input pattern at equilibrium. As a consequence, that value $g(x_0)$ can easily be compared with a reference value at node r to decide if it is necessary to create a new category. The output function of node r, $h(r)$ is a threshold function that indicates whether a new category need be created. When its output is 1, a new category has to be created. When its output is 0, the best–matching category should be adjusted instead.

The creation of a new category progresses in three steps. First, the recruitment network selects a previously unused node for the new category. In the next step, this node receives a learning signal from the recruitment network which makes it learn the current input pattern. In the last step, the newly–formed category node disconnects itself from the recruitment system to consolidate the category.

The recruitment network consists of another competitive network with one node y_i for each node x_i in the category network. Every recruitment node receives the recruitment signal $h(r)$ through an initially random weight v_i. This means that when $h(r)$ signals that a new category should be formed, all nodes y_i will receive a random activation. Since the recruitment network is competitive, the node with the largest input will inhibit all the other nodes and will then be able to create a new category in the network above.

The creation of the new category is controlled by the learning signal $f(y_i)$ sent from the winning node to its corresponding category node x_i. As the weights change toward the input signals, the activity of x_i will gradually increase and will consequently make the output of $g(x_0)$ larger. When x_i responds sufficiently well to the input pattern, $g(x_0)$ will be large enough to shut off the recruitment system by inhibiting the node r. When this happens, learning in the category system will stop.

The learning connections from the x_i nodes to the y_i nodes will be disinhibited and will start the disconnection of the recruitment system. Since the winning node x_m is the only active one in the category system, only the node y_m will receive a learning signal. When this happens, the weight v_m decreases until the recruitment node y_m can no longer be activated by the $h(r)$ signal. When v_i=0, the category represented at x_i will have been consolidated. It cannot be changed by the recruitment system any longer. A complete account of the calculations in this network is given in appendix F.

There are some ideas incorporated in the above network that deserve some comment. One is the disconnection of the learning mechanism as learning progresses. Such a mechanism has been proposed as a model of the projection between hippocampus and entorhinal cortex (Rolls 1990). The hippocampus is assumed to operate as a learning mechanism that reads in new categories to cortex. When the categories are created, hippocampus is disconnected, which makes the new categories immune to subsequent change.

Another comment that can be made of the network is that is uses unlearning, or decrease of connection strength, as an important part of its operation. Although no other comparison is intended, this is a process that is known to operate in the Purkinje cells of the cerebellum (Ito 1982). When they change their conductivity, it is by a process of long term depression, that is, they decrease their ability to let signals through.

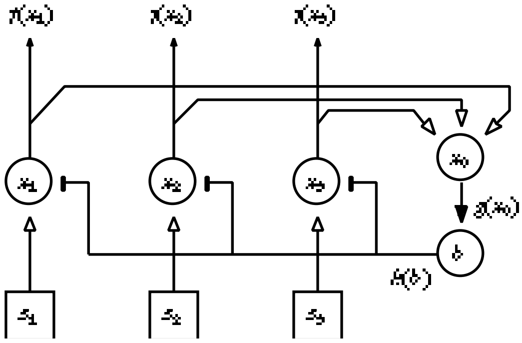


Figure 7.2.3 A network that normalizes its input pattern. The node x_0 receives the magnitude of the input vector $s_1 \ldots s_3$ and adjusts the bias b until the input is normalized. A normalized input vector is necessary for successful competitive learning.

One further important property of the perceptual learning system is that it does not need any external reinforcement signal. The learning of new categories is controlled by the input signals on their own. There is, however, the possibility that the matching criterion is temporarily altered in order to create new categories, as will be described below.

Finally, we note that the output of node $h(r)$ is a representation of the familiarity of a stimulus. If it is known, the output of $h(r)$ is close to 0, otherwise it is close to 1. This signal is, thus, a good candidate for the activation of an exploratory behavior. In section 7.4, we will see how the output of a familiarity detector can be used to increase the exploratory motivation of the animal.

The learning equations above work best when the input pattern is normalized, that is, when $\Sigma_i s_i=1$. This can be accomplished by adding yet another level of processing. Such a network is shown in figure 7.2.3. It iteratively computes the normalization factor for the input pattern and then facilitates the input signals appropriately. The sum of the input signals is computed at the node x_0. This node has a threshold of 1, which means that its output will reflect whether the sum of the inputs is larger or smaller than one. If it is larger, the node b will become inhibited, if it is smaller, the output of d will decrease instead. Since d integrates its input over time, this process will continue until the weighted sum equals one, that is,

$$\sum_{i=1}^{n} h(b) f(x_i) = 1. \tag{7.2.1}$$

We can compare this initial scaling of the input signals with the processes that are assumed to take place in the thalamus. Although they are much more complex than a simple normalization, one role of the thalamus appears to be to limit the dynamic range of its input (Mason and Kandel 1991). This function is closely related to normalization (Grossberg 1976). Also compare this processing step to the idea of a sensory buffer or sensory register (Sperling 1960). Appendix G describes this normalization process in detail and also shows how the network can be changed to use the Euclidean norm.

Categorical Learning Based on Expectations

With the learning mechanisms presented above, we can now investigate how the perceptual system interacts with the learning mechanisms described in chapter 5. Figure 7.2.4 shows the overall architecture of the learning system of our creature. The normalization module will limit the range of the sensory input signals, which will then continue to two later processing stages. One is the categorization mechanism presented above. This system produces categories that are used as input to the associative learning system, which in turn generates expectations on future stimuli and conditioned behavior. The other is a matching module that compares the perceived categories with the expected ones. As we saw in chapter 5, this is a general principle involved in many types of learning.

The matching between expected and actual stimuli serves two purposes in this system. The first is to send a learning signal to the categorization mechanism. This signal will activate the node *r* in the categorization network shown in figure 7.2.2, which in turn will start the formation of a new category. The second function of the matching module is to activate an orienting behavior toward the unexpected stimulus. This process requires that any ongoing behavior is first inhibited through a behavioral inhibition system like the one described in section 4.2. If the orienting behavior module generates both orientation and approach behaviors, the system in figure 7.2.4 will implement an efficient exploratory behavior. As long as some stimulus in the environment is unexpected, the creature will continue to explore it by alternating between orientation and approach behaviors. Simulations of this exploratory mechanism will be presented below in section 7.4.

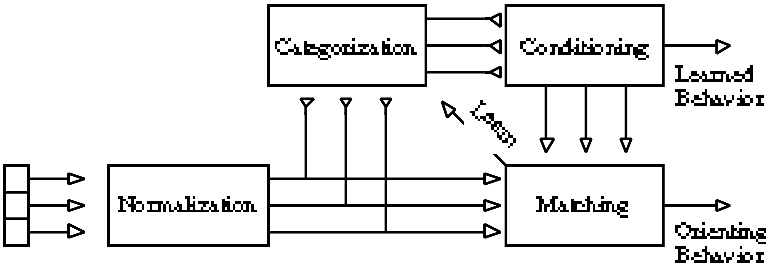


Figure 7.2.4 The interaction between normalization, categorization, conditioned expectations and matching. When a mismatch occurs between expectations and the current input, the ongoing conditioned behavior is inhibited, an orienting reaction is executed and the learning system is activated. This will let the creature learn about the unexpected situation or stimulus.

We can again compare the function of the matching–recruitment system with some theories about the function of the hippocampal system. It has been suggested that the role of the hippocampus is to create orthogonal representations for the activity patterns it receives from entorhinal cortex. These representations subsequently act as tutors for the cortical neurons and let them learn categories that are otherwise not linearly separable (Rolls 1990).

It is likely that the subiculum of the hippocampal formation compares expectations with the actual situation in a way that resembles the matching system proposed here (Gray 1995). We may note in passing that the matching system also plays a third role that was discussed in section 5.7. This is to select information for storage in what we may call a “working memory”. This mechanism was used in section 5.7 to store memories for new types of food for subsequent correlating with sickness. This is again a function that has been ascribed to the hippocampal system (Lynch and Granger 1991). We will return to this mechanism in chapter 9 and discuss its role as a more general working memory.

Figure 7.2.5 shows a simulation of the system in figure 7.2.4, where the creature was faced with the so called XOR problem (Minsky and Papert 1988). As described in section 3.5, this is the simplest example of negative patterning. First CS_1 and CS_2 are independently associated with the the US by classical conditioning. In the next phase, the compund stimulus $CS_1 \bullet CS_2$ is shown and paired with no–US. This will cause severe mismatch to occur in the matching module, that is, in the d nodes, which will recruit a new category in the categorization system. Once this category has been created, it can be associated with the omission of the US and override the predictions from the individual cues. If the predictions from the individual cues should not be extinguished, the output from the category node must be sufficiently larger than the individual associations.

Note that a mismatch would have occurred even if the compound had been followed by the US since the predicted level would be twice the actual level. In this case, the mismatch would be of opposite sign. This situation is called *overexpectation* (Rescorla and Wagner 1972).

The network presented here is similar to the ART network since new categories are created immediately when a mismatch occurs (Carpenter and Grossberg 1986). An alternative solution is used in the backpropagation procedure where weights are adjusted gradually until the required category nodes have been created (Rumelhart, Hinton and Williams 1986). This is also the approach taken by the model of classical conditioning developed by Schmajuk and DiCarlo (1992).

7.3 Place Approach

We have so far assumed that all stimuli of interest are perceptible from the position where the approach behavior starts. When this is the case, stimulus–approach behavior is sufficient. It is well known, however, that most animals does not use this strategy in many cases. Even when the goal is readily perceptible from the start location, animals often make use of a *place–approach* strategy (see section 2.7) or *piloting* (Gallistel 1990). In this section we will develop a very simple mechanism for the learning of this type of behavior.

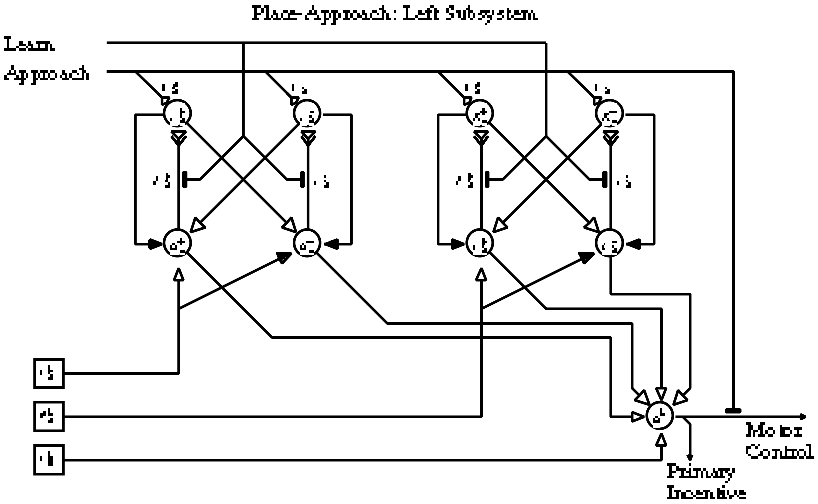


Figure 7.3.1 A behavior module for simple place–approach. Matching of the input pattern against a place prototype guides locomotion toward a specific place, even if the stimulus at that place is missing. Note that the reinforcement module from section 5.2 is the main component in this system.

The idea behind the place–approach system described here is that instead of using a single stimulus to guide locomotion, the whole stimulus situation is used. The goal is represented, not as a single odor, but as the whole smell complex of the goal object. Instead of comparing the intensity of the smell at the left and the right sensors, the place–approach system compares the whole stimulus situation to the left and the right with the reference situation at the goal. Given that the creature is already sufficiently close to the goal, the difference between the left and the right match can be used to guide the creature toward the goal. See also appendix H.

Figure 7.3.1 shows a simple behavior module for place–approach. It is assumed that the goal location can be recognized by the smell received at s_0. When this smell is present, it can directly guide the approach to the goal location. When the creature reaches the goal, the *Learn* signal is assumed to become active. This will gate the learning process in the reinforcement modules, which will consequently learn the configuration of stimuli at the goal location. In the figure, this configuration consists of the signals from only two sensors s_1 and s_2. To predict the location of the goal accurately, three or more stimuli are needed. This *prototype* for the goal, represented by s_0, s_1 and s_2, is stored in the weights w_i in the reinforcement modules. The circuit described here, thus, implements a detector that is tuned to a specific stimulus configuration (compare Schmajuk 1990).

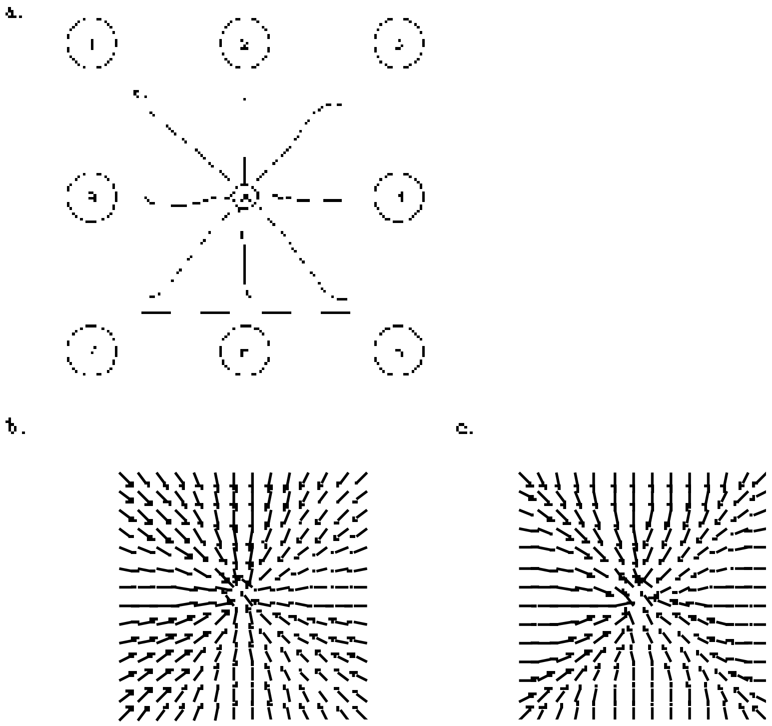


Figure 7.3.2 Place approach using a configurational stimulus. (a) The path taken toward the goal, A, from different start locations when the creature has learned the location of the goal. Note that the goal does not give off any smell in this simulation. The stimuli 1–8 are used to locate the goal. (b) The life-–space generated by the goal on its own when it does smell. (c) The life–space generated by the stimuli around the goal. This is a fairly good approximation of the life–space in (b).

At later times, the behavior module can operate in two modes. In the first, it uses the sensation of the goal directly, as above, and approaches it using a stimulus–approach strategy. In the second mode, the complete sensory situation is compared with the prototype pattern stored in the reinforcement modules. The match between the prototype and the actual sensory situation is subsequently used to control the approach behavior. Since the behavior module is divided into one left and one right half, approach can be accomplished in a way analogous to that for stimulus approach. Figure 7.3.2 shows the path taken by a creature as it approaches a hidden goal location using a configurational category. Note that the network in figure 7.3.1 does not make the creature take the shortest path to the goal. The selected path has the property that the sensors on the left and right of the creature match the stimulus configuration equally well. To approach the goal on the shortest path, a more advanced learning mechanism is required. As can be seen, the configuration stimulus predicts the path to the goal fairly well.

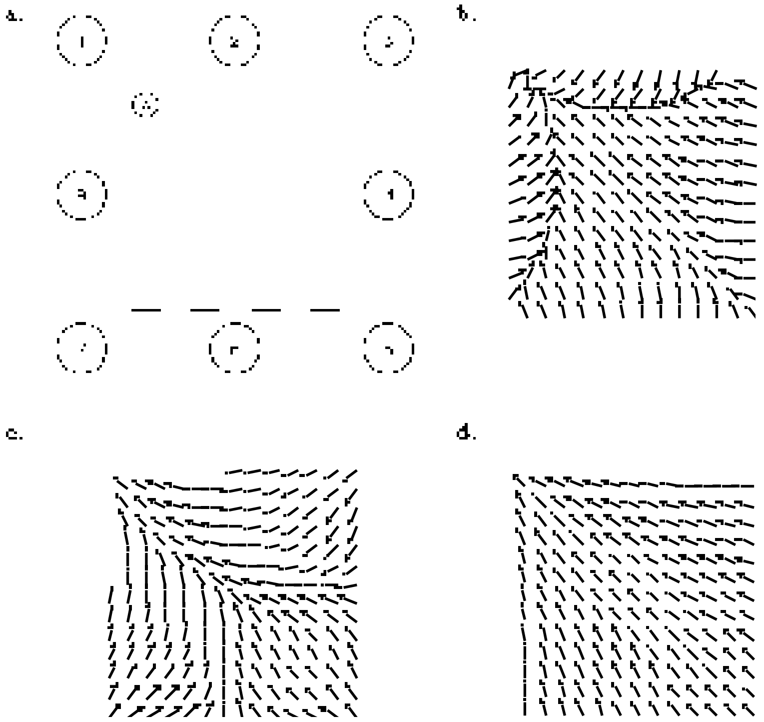


Figure 7.3.3 A creature approaching a goal at a corner of the environment. (a) The environment. (b) The life–space generated by the goal itself mixed with the configurational representation of its location. (c) The life–space generated solely by the configurational stimulus. (d) The life–space created by the scent of the goal without the aid of a configurational stimulus.

In figure 7.3.3, the creature has learned a goal location that is not in the center of the environment. In this case, the life–space set up by the configurational stimulus is less accurate. However, it is still sufficient in order to guide the creature to the goal.

In general, the stimuli guiding the approach of the goal should be selected to be far away from the goal itself. This is obviously not possible for a simple modality like olfaction. In real animals, vision is used for this task instead since it easily selects *distal landmarks* instead of local cues for place approach behaviors. Many models exist for these processes (for example, Zipser 1985, Prescott and Mayhew 1993, Schmajuk and Blair 1993, Touretzky and Redish 1995). However a closer examination of these models is beyond the scope of the current presentation.

An interesting aspect of the place–approach module presented here is that it uses two symmetrical systems to guide locomotion. This can be compared with the model developed by Schmajuk and Blair (1993). In their model, only a single value is generated as output. This requires that some external system try out different locations to estimate the gradient around the goal. An alternative approach was used by Barto and Sutton (1982, see also Barto, Anderson and Sutton 1982) who used a single output as a pay–off signal that reinforced the appropriate responses. When two symmetrically organized systems are used as in the present model, the relation between two spatially displaced locations is used to guide locomotion directly (see section 4.2).

To summarize, a behavior module that can learn configurational stimuli has two ways of guiding the creature toward a goal. The first is by a stimulus approach strategy that is used when the goal is perceived from the location of the creature. The second possibility is to match the configurational or situation representation of the goal with the present perception and decide on which way to move. In most cases, these two processes cooperate with each other.

7.4 Exploration

By combining the expectancy network presented in section 5.10 and the behavior module from the previous section, an interesting exploratory behavior can be generated. The idea behind this system is to let the creature orient toward all stimuli that are in some way unexpected. We have already seen how the required signals are calculated in the reinforcement modules of the expectancy network. When a stimulus is unexpected, the corresponding behavior module is activated, and the stimulus will be oriented to and possibly approached.

Figure 7.4.1 shows how the expectancy network can be used to activate the orienting system and cause the creature to approach an unexpected stimulus. When the creature is placed in a novel environment, its orienting mechanism will immediately become triggered and the creature will start to approach the closest stimulus using the behavior module described above[3]. This behavior will continue until the stimulus becomes expected in terms of the other perceptible stimuli. When this happens, the orienting system is temporarily habituated and the next unexpected stimulus receives attention.See also appendix I.

Figure 7.4.2 shows a simulation of this combined network. In this simulation, eight behavior modules were connected to an expectancy network with eight reinforcement modules. The complete network uses 568 nodes[4]. As can be seen, this results in a sequence of approach behaviors and learning until the entire environment has been explored (figure 7.4.2a). Initially, all stimuli in the environment attract the attention of the creature. This life–space is a picture of unrestrained curiosity (figure 7.4.2b). When the exploration phase has finished after 10,000 time

[3] For simplicity, it is assumed here that the accelerating approach behavior generated by the place approach system is used. It would, however, be more accurate to use a combination of orienting and a combined approach behavior as described in section 4.2.

steps, however, the behavioral inhibition is withdrawn and the creature can continue its original business, in this case, aimless wandering (figure 7.4.2c). The various stimuli attract the creature very little (figure 7.4.2d). The vectors around stimulus 7 are artefacts generated by the the impoverished stimulus situation. When stimulus 3 disappears behind the wall, the expectation of stimulus 7 decreases temporariliy and is classified as novel. This dependence on all the available stimuli decreases with an increasing number of stimuli in the environment.

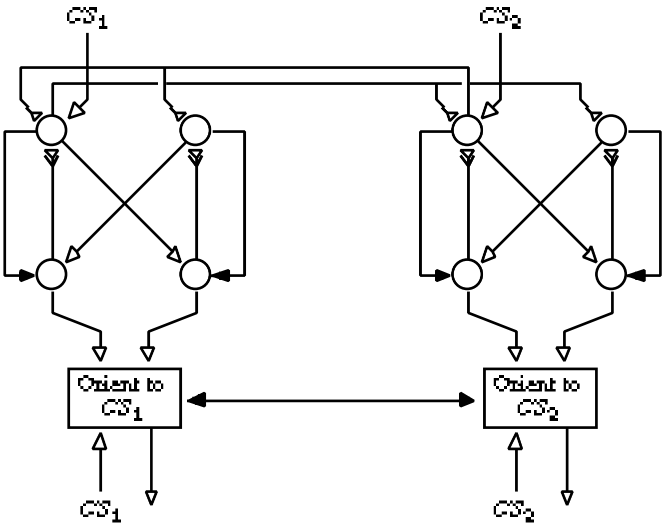


Figure 7.4.1 A network that generates exploratory behavior through the interaction between a habituation system, the expectancy network and the orienting reaction. The reinforcement signals generated when unexpected (or omitted) stimuli are encountered activate an orienting reaction toward that stimulus. As a consequence, the creature will approach each unexpected stimulus in turn and learn about the environment until all stimuli become expected.

In figure 7.4.2d and e, the stimulus 7 was moved from one part of the environment to the other. Since the omission of stimulus 7 is unexpected, the behavioral inhibition will become activated and stop ongoing behavior while simultaneously starting an orienting reaction. The nodes that react when an expected stimulus is omitted can be likened with the misplace cells described by O’Keefe and Nadel (1978).

4. The creature can recognize 8 smells. This requires place–approach modules with eight reinforcement modules each, that is, 8×4 nodes. Since the behavior module is divided into two symmetrical halves, twice as many nodes are necessary. To this we add the two approach nodes that generate the output from each behavior module. This results in 2×(8×4+1) nodes per behavior module. To this, we add the five nodes in each reinforcement module of the expectancy net, giving a total of 71 nodes per smell. Since eight smells are possible, this number must be multiplied by eight we and end up with 568.

This mechanism is very important since it allows the creature to react to stimuli that are not present in its environment.

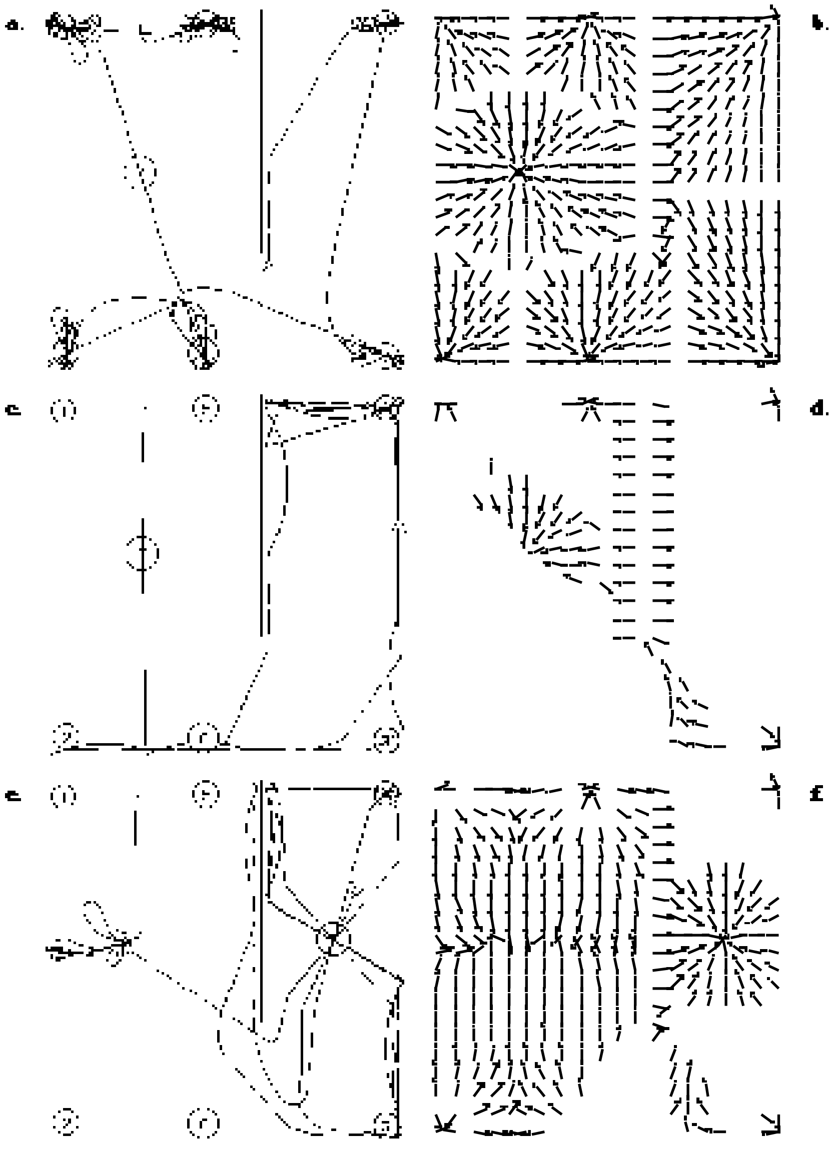


Figure 7.4.2 Exploring a novel environment. (a) The initial behavior of the crea-

ture. (b) The initial life–space. (c) Behavior after 10,000 time steps when the orienting reaction has habituated. (d) The new life–space. (e) Behavior when object 7 has been moved. (f) The corresponding life space.

As can be seen, the creature first approaches the location where the object used to be by using the piloting mechanism described in the previous section. This behavior includes a searching phase during which the creature tries to find stimulus 7 in the vicinity of its original location. This is also the case for many real animals (Bell 1991, Gallistel 1990, Ellen and Thinus–Blanc 1980). For example, if a desert ant does not find its nest in the expected location, it will start to search the area around the expected location until the nest is found[5] (Gallistel 1990).

When the orienting system becomes habituated, the creature will start to wander again. After a while, it will perceive the stimulus 7 in a new location, which will again trigger the exploratory behavior.

7.5 The Valences of Percepts – Affordances

In the view of perception we have presented above, a situation or object is always perceived in terms of its valence as desirable or undesirable at a certain moment. These valences are represented as vectors in the life–space of the creature. In figure 3.4.2, all objects were initially appetitive as a result of their novelty. As the creature explored the environment, the different stimuli became less and less attractive until they finally vanished completely. Following the suggestion in chapter 4 that objects are perceived when they can be acted on, we may consider the stimuli in the environment essentially invisible after the exploratory phase. If the motivational state of the creature changes, the stimuli may change valences and become perceptible again. Stimuli are thus perceived only when they constitute a potential goal or subgoal for a particular motivation.

This view of perception has many parallels to Gibson's theory of affordances (Gibson 1979). In this theory, objects in the environment are considered to afford different actions. Food affords eating, water affords drinking, a path offers locomotion, etc. In fact, Gibson claims that perception *is* the pickup of affordances. The affordance of an object is considered as an invariant of the environment. This means that it is independent of any of the needs and dispositions of an observer.

> The concept of affordance is derived from these concepts of valence, invitation and demand but with a crucial difference. The affordance of something does *not change* as the need of the observer changes. The observer may or

---

[5] Since the ant's nest is not supposed to be moved, the ant does not have any mechanism for termination of its search behavior. If the nest really has been moved, or the landmarks used to guide the animal have changed, the ant will search until it dies of hunger (Gallistel 1990).

> may not perceive or attend to the affordance, according to his needs, but the affordance, being invariant, is always there to be perceived. (Gibson 1979)

This is one point where our view differs from that of Gibson and is more in line with Lewin's original use of the concept (Lewin 1935). In his view, affordances were identical to valences, which of course, are always related to the needs of the creature. As we have already argued, it is also necessary to add a novelty–familiarity dimension to the hedonic values to recognize which stimuli afford exploration.

An alternative, and more common, way to define perception is to consider a stimulus as perceived when it is somehow represented within the sensory system of an animal. Such a view of perception is problematic in that there is no way to figure out whether an animal has perceived a stimulus or not. Placing the discussion of perception more on the sensory side makes the inclusion of attentional mechanisms more natural however.

7.6 Attention

With the mechanisms we have presented so far, it is very easy to include attention in the creature. We simply let the current drive state influence the perceptual processes. How this can be done is shown in figure 7.6.1. The drive state is allowed to bias the perceptual processes at the behavior modules B_0 and B_1 by a facilitating connection. As a result, the creature will more easily perceive stimuli that are relevant for its current needs.

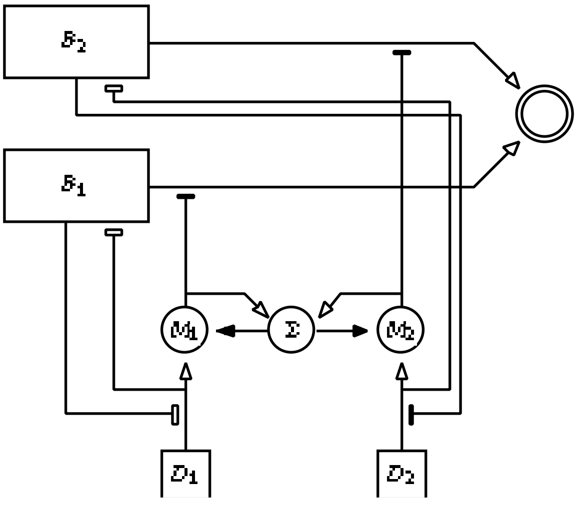


Figure 7.6.1 Attention is included in the creature by biasing the perceptual processes by the current drive state.

called *recurrent expectations*. This process will continue until the goal representation, *p*, is activated. When this happens, the node *p* activates the two nodes a_1 and a_2. The plastic connection u_1 will consequently sample the activity level of the goal representation at a_1. The connection u_1 is used as a short-term memory store of the goal prediction for CS_1. The output of *p* will also reset the node v_1 which will stop further sampling at u_1. The processing of stimulus CS_2 will then proceed in the same way.

Because of the discount factor used in the learning process, the activation of the goal representation will be weaker with an increasing number of associative steps. As a consequence, the activation of the goal representation will be strongest for the stimulus with the smallest psychological distance to the goal. The weights on the connections u_1 and u_2 can, thus, be considered as representations of how well CS_1 and CS_2 predict the goal.

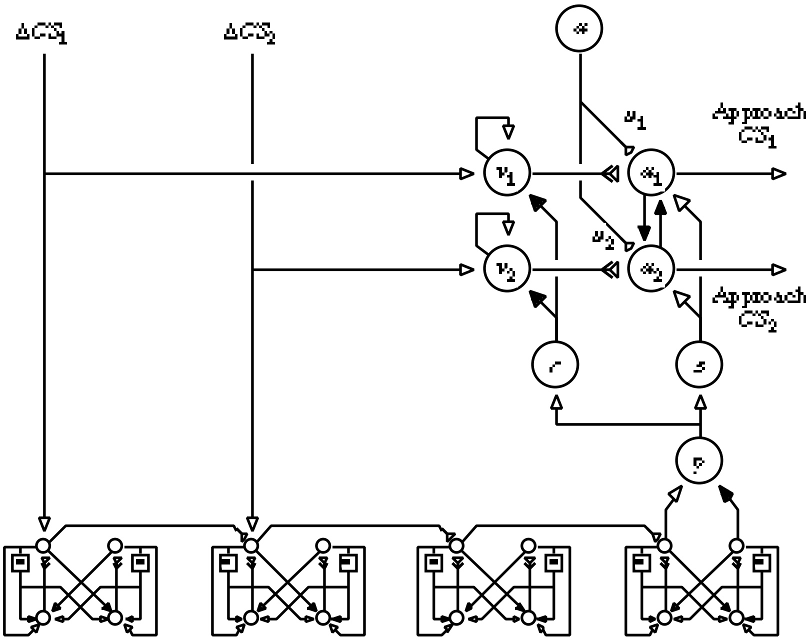


Figure 8.4.1 Using the expectancy network for calculation of goal prediction. In the simulations described in the text, eight stimuli and approach behaviors were used. In this case, the total network uses 52 nodes. See the text for further explanation.

By activating node *a*, the activity of the nodes a_1 and a_2 will reflect the goal predictions of the two stimuli. Since the two nodes a_1 and a_2 compete with each other as described in section 3.3, only the node corresponding to the best goal predictor will emit a signal. This signal will be used in the second phase to approach the best goal predictor. Since the creature uses its expectancies to look ahead along the evaluated path, this type of subgoal selection will be called *look–ahead choice*.

A related function of motivation is to bias the selection of the motivational state in yet another way. In chapter 6 we saw that it was necessary to select one engagement at a time if behaviors are not to interfere with each other. Such a choice has a cost, however. Stimuli that are not relevant to the current engagements become invisible (See figure 6.3.2). When attention is introduced, this problem is automatically solved.

The creature can switch attention between different stimuli instead of blending the behaviors produced by them. The creature will consequently engage in only one behavior at a time, but will still be able to perceive all stimuli. Its behavior will, thus, be much more flexible.

7.7 Imagination

There is one aspect of perception that has been deliberately left out in the presentation so far: that is its ability to construct complex percepts by top–down processes (Shepard 1984). While this mechanism is very important, for example, in vision, it can only play a very limited role in the olfaction used by our artificial creatures. Here, we will only sketch three possible ways in which ‘imagination’ could be incorporated in the system presented in section 7.2 (See figure 7.7.1). A more thorough analysis would require a book of its own.

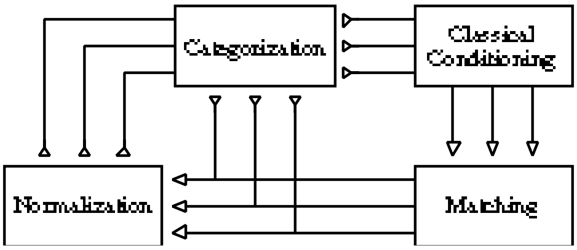


Figure 7.7.1 Three possible pathways for imagination and active perceptual reconstruction.

The first mechanism uses a pathway from the categorization system to the normalization module to fill in a recognized category. The result of this process is to make the stimlus representation more similar to the prototype of the best category than is in fact the case. A mechanism of this type is used in within the adaptive resonance theory to stabilize the learning process in a competitive network (Carpenter and Grossberg 1986). Note that the associations from categories to sensations are atemporal, that is, they are read out immediately and blended with the input signals. This is, thus, a type of $S(t)$-$S'(t)$ association mediated by categories (See section 2.7).

A connection from the learned expectations in the conditioning module to the categorization system would make it possible for the creature to perceive its expected categories as if they were externally generated. This type of imagination differs from the one above in that it is categories which are imagined, not sensory data. In both cases, perception can be seen as “externally guided hallucinations” (Shepard 1984). A neural network model which is in agreement with this idea was presented by Trehub (1991).

If there is a time delay between the readouts of each expectation, it will be possible to generate sequences of expectations. Each category generates expectations of other categories, that in turn generate even more expectations. A temporal read–out is natural since classical conditioning is based on expectations of the future. In other words, we have a case of $S(t)$-$S'(t+1)$ associations. In chapter 7, we will argue that this mechanism is the basis for planning and higher cognition.

Is it possible to use the same mechanism for both $S(t)$-$S'(t)$ and $S(t)$-$S'(t+1)$ associations? A natural extension of the system above would be to let the expectation system learn $S(t)$-$S'(t+\Delta t)$ associations where the time delay between the ‘premise’ S and the expectation S' could vary. Associations of this type would generate expectations of the type, ‘when S occurs, S' will occur after a time delay of Δt’. Some models have tried to develop mechanisms of this type. See, for example, Gallistel (1990) and Schmajuk (1990).

A final imagination mechanism could be constructed by letting expectations which match sensations sufficiently well strengthen those sensations. A pathway from the matching module to the sensory register could mediate such a mechanism. This mechanism could possibly be used to incorporate expected categories in new categories. Can this be used as a basis for a chunking mechanism? This is a question we will address in section 9.4.

All three mechanisms have in common that they let the animal perceive its environment in terms of its own categories and expectations. When mechanisms of this type are included in a creature, we have something that can be called both *bottom up* and top *down processes* (Glass and Holyoak 1986).

7.8 Conclusion

Perception rests on two interacting processes: categorization and association. In our artificial creature, categorization is handled by a competitive network with recruitment learning. This system creates new categories when the already established do not represent the input pattern sufficiently well. Associations are handled by a conditioning system that learns associations between categories. These associations are treated as expectations that are matched against the input pattern. When expectations are not met, learning in the categorization system will be triggered.

An important property of perceptual learning is that external reinforcement is not necessary. Learning is governed by how well the internal categories and expectations match the external input. No reward or punishment is necessary. However, to produce *behavior* that uses the perceptual categories, reinforcement may be needed (compare section 2.5 and 8.3).

We have seen how top–down and bottom–up processes can be combined at a number of places in the system. Exploratory behavior was seen as a consequence of orienting behavior toward unexpected stimuli. Such a mechanism will automatically make the creature explore the environment as long as some perceptible stimulus is unexpected. Another important aspect of perception is that it interacts continuously with both motivational and behavioral processes.

Finally, we repeat the observation that it is not possible to separate perception from other processes on physical grounds. There is no specific part of the nervous system responsible for perception and some other that is not. The different systems involved in perception are also used for behavioral learning and motivation.

Chapter 8

Spatial Orientation

8.1 Introduction

In the previous chapters, the various learning mechanisms have only been used for very simple problems. In this chapter we will see how an artificial creature can use the mechanisms we have presented above to solve a number of problems in the spatial domain. We will investigate two different types of learning mechanisms. The first is procedural learning which uses the reinforcement module presented in chapter 5 to learn sequences of behaviors. The second mechanism uses the expectancy network as its main component to represent knowledge about the environment, and uses it for look–ahead choice.

It is important to distinguish between the *orienting reaction* described in sections 4.4 and 9.2 and the concept of *spatial orientation*. The orienting reaction directs the sensory apparatus of an animal toward a stimulus in the immediate environment, while spatial orientation refers to an ability to orient from one place to another. A possible way to avoid confusion would be to call the present chapter spatial navigation. Since spatial navigation usually refers to more advanced mechanisms than the ones which will be described here, we have chosen the first term, however.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

8.2 Innate Behaviors

The creature will have four types of innate behavior modules. The complexity of these modules varies considerably depending on the number of stimuli that are used to control the behavior, and on whether they include learning or not (figure 8.2.1).

The first kind is a set of primitive *responses*, that is, behavior modules which do not use feedback and, thus, cannot be goal-directed. The primitive response repertoire of the creature will generate behaviors such as *move–ahead*, *turn–left*, *turn–right* and so on. At least in theory, all behaviors can be composed of these primitive responses, but as we will see below, this is not a very efficient way to learn complex behaviors. We will also include systems for activities such as grooming within this set of behavior modules. The common denominator for these modules is that they do not use sensory input to control the generated behavior once it has started. In figure 8.2.1, an optional stimulus is shown with the response generation behavior module. This is intended to illustrate that a sign-stimulus can activate the module, but does not control the execution of the behavior.

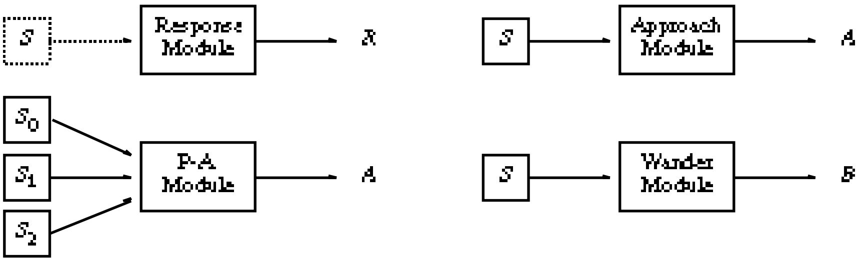


Figure 8.2.1 Four types of innate behavior modules.

The behavior modules of the second type generate *stimulus–approach* behaviors. These modules have the internal architecture shown in figure 4.2.10. Each module is, thus, constructed for one specific stimulus. When no learning is used, it is necessary that the creature has one behavior module for each potential stimulus type in the environment.

The behavior modules of the third kind control *place–approach* behavior (see sections 2.7 and 7.3). These behavior modules must naturally include some learning ability since the stimulus situations at interesting places cannot be innately known. However, given the learning abilities we have presented above, the behavior module, as such, must be innate. The approach module presented in chapter 7 recognized one stimulus with a specific meaning. This stimulus played the role of an unconditioned stimulus, and could be approached on its own. The other stimuli were used to predict the location of this special stimulus. An approach system of this type includes both a place–approach and a stimulus–approach module.

Finally, the creature needs wandering and obstacle avoidance behaviors which help it move around in the world as described in section 4.4. They are used to follow walls and corridors, to avoid obstacles and sometimes for random walk.

8.3 Procedural Learning

In procedural learning, the creature learns what it should do in each situation to accomplish a certain goal. This ability requires that the creature can coordinate behaviors in two ways. The first is to associate the activation of the appropriate behavior module with each stimulus context it encounters on its way to the goal. The second type of coordination concerns the order in which behavior modules should be activated. In many cases, either one of these mechanisms is sufficient.

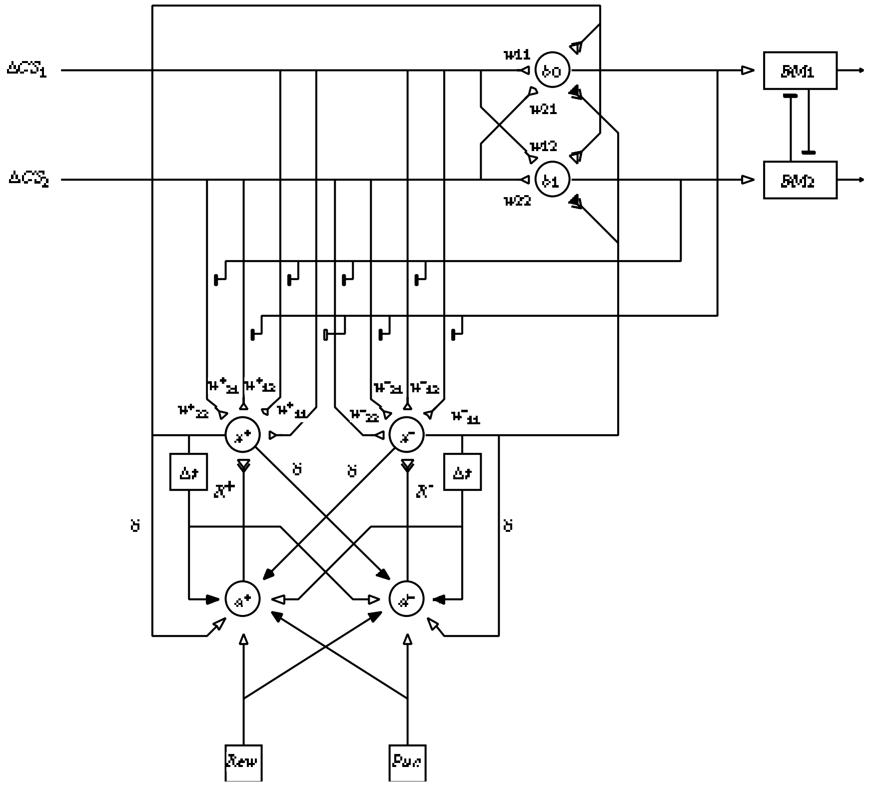


Figure 8.3.1 The general layout of a sequential learning system. The sensory cues CS_1 and CS_2 are associated with two behavior modules BM_1 and BM_2, by the reinforcement module. After learning, the system will activate the behavior modules according to the discounted reward it expects to receive if the corresponding behavior is performed. In the simulations described in the text, four behavior modules were used and not two. In the larger simulations, 512 stimuli were used which required 4096 facilitated connections (only 8 are drawn here) and 5120 plastic weights (12 are drawn here). The activity of the nodes b_i are assumed to reflect the selected behavior.

In the simulations to be presented below, we will investigate two types of procedural learning. The first will be based solely on responses, and will, thus, be an example of stimulus–response learning. The second will use only stimulus–approach modules. Figure 8.3.1 shows the general layout of these learning systems.

Depending on the type of learning required, the behavior modules can be any of the four kinds described above. In a more advanced creature, all types of behavior modules interact with each other. In either case, it is necessary to select among the different behaviors generated by the behavior modules since the learning system does not guarantee that only one behavior module at a time is active. In our first simulations below, arbitration by probabilistic behavior selection will be used as described in section 4.3. Using this arbitration scheme, the probability of a certain behavior is proportional to the activation received by the behavior module. See appendix E for a formal presentation of the network. In the next few sections, we will see how a creature behaves using different types of procedural learning.

Stimulus–Response Chains on a Grid

The first simulation is intended to show that the learning system presented in chapter 5 is able to learn an optimal behavior. A creature is equipped with four basic responses and a procedural learning system of the kind described above. The four responses will be called *move–north*, *move–south*, *move–west*, and *move–east*. These behaviors simply moves the creature a fixed distance in either direction in relation to the environment.

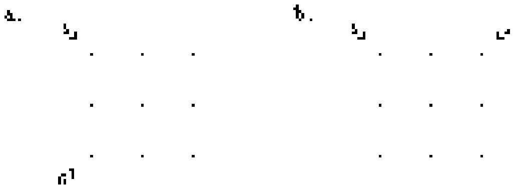


Figure 8.3.2 A simple environment where the creature can occupy only 64 locations. The problem for the creature is to learn the shortest path from the start (S) to the goal (G). (a) The initial environment. (b) When the creature has learned the behavior sequence that leads to the goal, the goal is moved and the creature has to modify its behavior.

The sensory input consists of a number of situation categories that are generated by a a fixed grid (figure 8.3.2). In this first simulation the outcome of performing an action in a certain state will be entirely deterministic and moves the creature from one location on the grid to another.

The simulation is divided into a number of trials. At the beginning of each trial, the creature is placed at the start location (S) of the environment shown in figure 8.3.2a. The creature is then allowed to try out the different responses at random until it reaches the goal location (G). When the goal is reached, the creature performs its consummatory behavior and is again placed at the starting location. This procedure is repeated until the creature moves from start to goal in an efficient manner. The final level of performance will depend on the selection of actions. As described in section 4.3, a temperature parameter controls the level of randomness in the action selection. Since this choice is not deterministic, the creature will not always take the shortest path from start to goal.

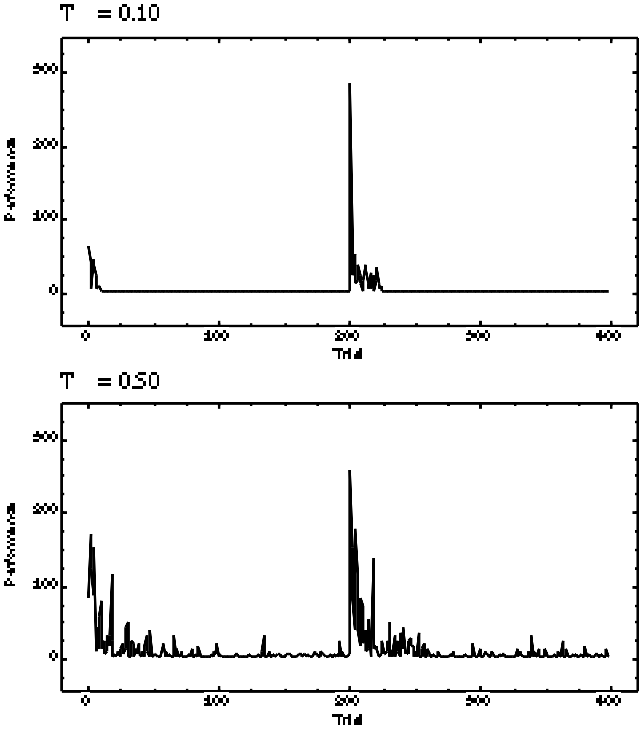


Figure 8.3.3 The performance in the 64 state environment shown in figure 8.3.2a. When the performance has stabilized after 200 trials, the goal is moved and the creature is allowed to relearn its response sequence. The simulation was run with a temperature (T) of 0.10 and 0.50. With random behavior, it takes on the average 51 steps from start to goal. Note that the creature performs much worse than random when the goal has been moved.

After 200 trials when the creature has learned the path from the starting point to the goal, the goal is moved as shown in figure 8.3.1b and the creature is allowed to continue its business in the environment. Since the creature must now relearn the action sequences leading to the goal, its performance will drop considerably. After 200 new trails, however, the performance has almost returned to the level established for the first goal location (Figure 8.3.3).

During the second phase of the simulation, the creature had a tendency to visit the previous goal-location at more or less regular intervals. The trials which required most steps to reach the goal in this phase nearly always started with the creature first moving to the last goal location, and then on to the new one. This behavior is a consequence of previously established associations, which have not yet been completely extinguished. From a biological perspective, behavior of this type is very sensible in many cases.

When responses were selected at random without any learning, the average number of steps from start to goal were 51. As can be seen in figure 8.3.3, the creature performs much worse than chance when the goal has been moved, and the environment has to be relearned. With a low temperature, relearning took much longer than the initial learning. With the higher temperature, initial learning and relearning were more similar. Since, the creature behaves more randomly at a higher temperature, it is more likely to deviate from the best path and consequently more likely to find the new location of the goal.

In general, the selection of temperature determines the relation between exploration and exploitation of the environment. It seems reasonable to use a high temperature when drive levels are not to high, and to decrease the temperature as much as possible when the drive level becomes higher. One role of the exploratory drive described in chapter 6 is to select a higher temperature for an engagement when it wins the motivational competition.

Stimulus–Response Chains

To make the situation more realistic, the creature was equipped with four responses, which can be called *move–ahead*, *move–slowly*, *turn–left* and *turn–right*. These behaviors send signals directly to the motors as summarized in table 8.3.4.

Response	m^L	m^R
move–ahead	1.0	1.0
move–slowly	0.2	0.2
turn–left	0.8	1.0
turn–right	1.0	0.8

Figure 8.3.4 The primitive responses in the second simulation.

These responses were chosen so that the creature could potentially end up in any location within the bounds of its environment. When the whiskers sensed a wall to the left or to the right, the creature was forced to select the action that turns away from the wall. A wall to the left would select the response *turn–right*, and a wall to the right would select the response *turn–left*.

Like in the previous simulation, the stimulus sensed by the creature was generated by a grid. Note that this grid was used to generate sensory information only, and not to control the movements of the creature. To let the creature know which way it was moving, the sensory state was also a function of the direction in which the creature was heading. This direction was represented with a resolution of eight directions (figure 8.3.5). All in all, this lets the creature recognize, 8×8×8=512 unique states, or stimuli. The creature is consequently not able to generalize its behavior across different locations.

When the creature is located in square i, j and facing direction d, the stimulus $CS_k=1$ for $k=8(i+8j)+d$ and $CS_k=0$ for all other stimuli. This should be compared with the infinite number of possible locations the creature can potentially occupy within its environment. The creature has, thus, only a very rough idea of where it finds itself at any time.

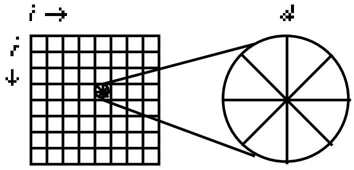


Figure 8.3.5 The sensory grid. The creature can recognize 64 different locations in the environment, and 8 different directions. Taken together, this results in 512 different stimuli.

Figure 8.3.6 shows how the performance of the creature becomes better for each trial at different temperatures. At a temperature of 0.005, the creature nearly always choose the best response learned so far for the current stimulus situations. As can be seen in the uppermost diagram in figure 8.3.6, this results in a decrease in the number of steps needed from start to goal for each trial. When the performance has reached a low level, however, a number of spikes appear in the performance level. These are a result of a rather narrow representation of the environment. Since the creature has nearly always chosen the response with the highest expected reward, it has only learned about the locations on the path from start to goal, and not about the other locations in the environment. If it, by chance, selects a response which leads away from the learned path, it will not be able to find its way back again. In these cases, its behavior will be essentially random which gives rise to the large spikes in its performance.

At a higher temperature of 0.010, the frequency of these spikes decreases and disappears completely after 450 trials (figure 8.3.6 middle). At the even higher temperature of 0.020, the spikes essentially disappear, but the performance becomes worse since the choice of response becomes less accurate (figure 8.3.6 bottom).

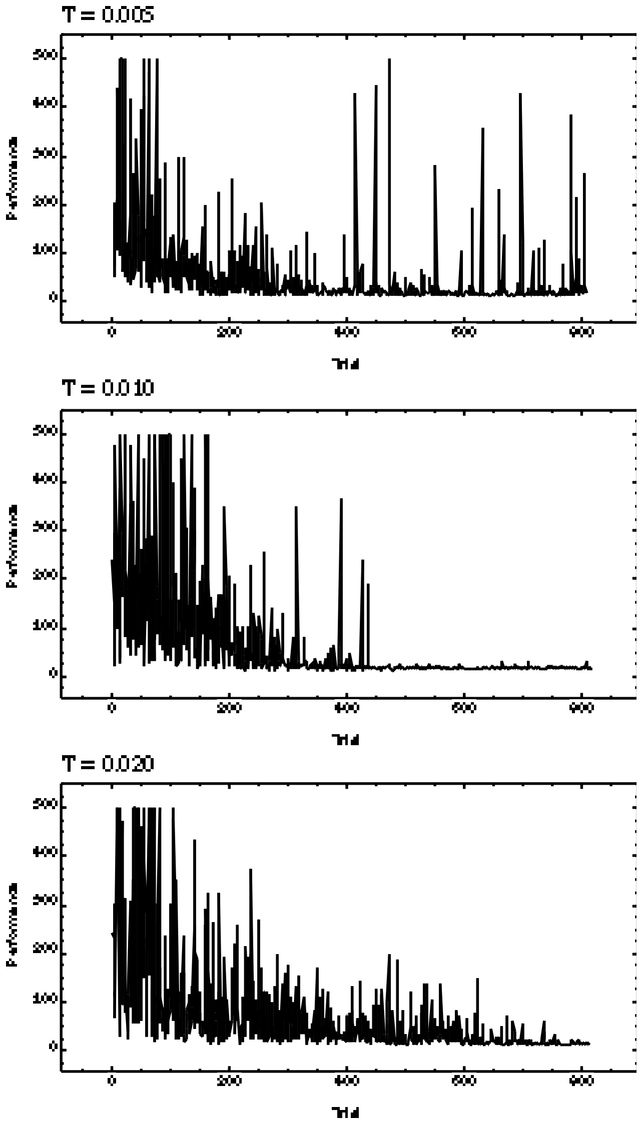


Figure 8.3.6 The performance in a 8×8×8 grid with the four responses, *move–ahead*, *move–slowly*, *turn–left*, *turn–right* with a stochastic response-selection and a temperature (T) of 0.005, 0.010 and 0.020.

In a final simulation, the behavior of the stimulus–response creature was tested in a simple maze. Figure 8.3.7 shows how the behavior improves over the trials. In the first trial, behavior is entirely, random and as can be seen, the creature spends a lot of its time moving back and fourth aimlessly. After 50 trials, the creature begins to perform more sensibly in the later half of the maze, but behavior is still random in the first half of the maze. With more trials, the behavior becomes better however. When the simulation was stopped after 200 trials, the behavior was nearly almost perfect. At times, the creature would deviate from the optimal path, however, as can be seen in the final illustration.

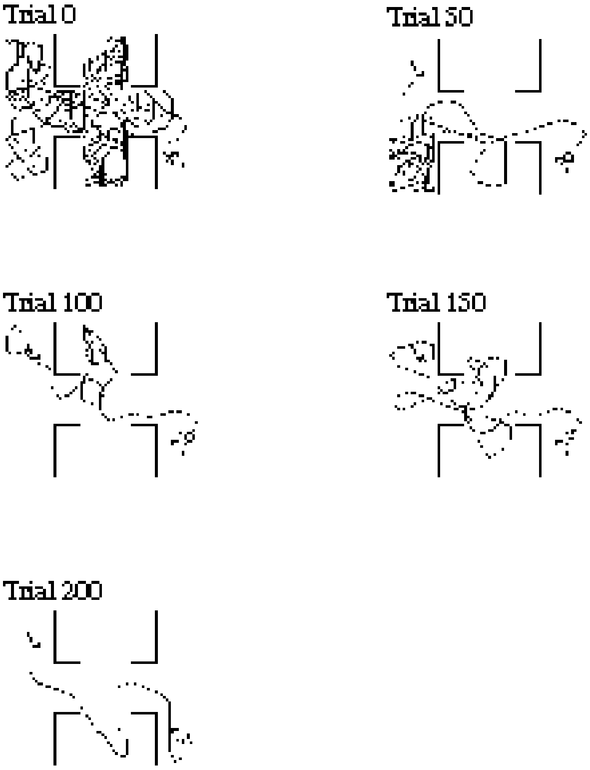


Figure 8.3.7 The behavior of the stimulus–response creature in a simple maze after different number of trials.

Stimulus–Approach Chains

The learning method described above can also be used with stimulus–approach behaviors. In this case, the stimulus–approach behaviors are substituted for the responses used above. A number of simulations were run to check how well the procedural learning system would handle chaining of approach behaviors. Figure 8.3.8 shows the learning curves from two of these.

In the first simulation, the creature had to learn the 8×8 environment described above. As can be seen in top of figure 8.3.8, the performance becomes better with training, although it is not clear whether it evetually converges to the optimal behavior or not. A number of simulations have been run in this environment, but none of them has converged on the optimal behavior. Note, however, that the behavior becomes much better than random search.

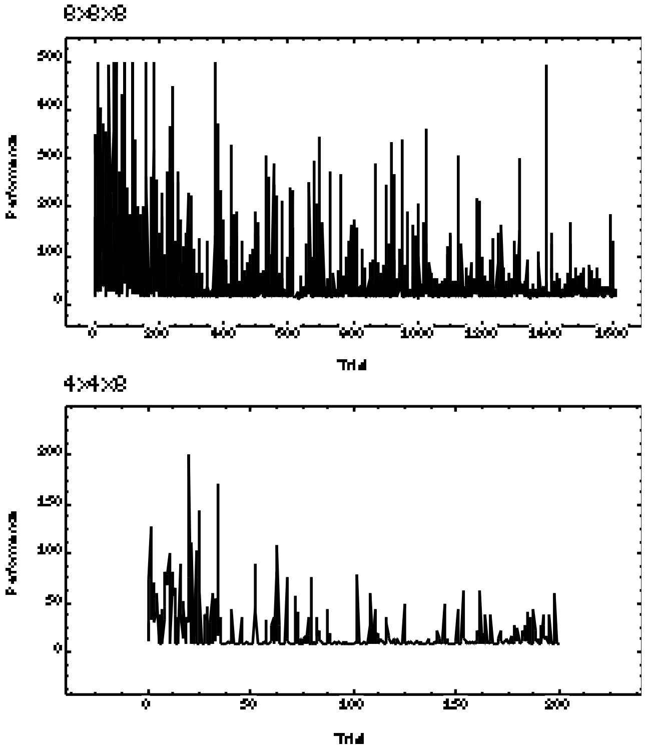


Figure 8.3.8 The performance when the creature learns stimulus–approach chains. Learning is very slow, although the performance becomes better with increased practice. It is not clear whether learning will eventually converge to the optimal level or not.

To test if the learned behavior could in fact converge, a number of smaller simulations were run in an environment that used only 16 grid locations. In this case, the learning process nearly always reached a nearly optimal level on most trials when the temperature was sufficiently low. There appears to be two reasons why learning was much better in the smaller environment.

The first reason is that fewer states had to be visited to learn the environment. This obviously leads to faster learning. This does not, however, explain why the learning did not appear to converge in the larger environment. We suspect that the failure of the learning in this case was a consequence of the continuous change of the selected approach behavior. This made it very likely that the creature would reach one stimulus while trying to approach another. As a result, the wrong behavior would very often be rewarded. In the smaller environment, the creature is more likely to reach the stimulus it tries to approach since, on the average, it is at a shorter distance. The incorrect behavior will consequently be less commonly rewarded.

If this suspicion is correct, it implies that a better stimulus–approach chaining could be obtained by forcing the creature to approach the selected stimulus for more than one time step before a new behavior is selected. This has not been tested in any simulation, however.

Dynamic Category Formation

In the above simulations, the stimuli used by the creature were generated by a fixed grid in the environment. It would be more realistic if these categories were generated internally by the creature, and not given explicitly. Recall from chapter 7 that there are two basic situations in which new categories are needed. In the first case, the best category does not match the input sufficiently well. In the second case, the expectations generated from the best category do not match the actual situation or reward. What will happen if these mechanisms are added to the chaining system described above?

A final simulation was set up to test if the creature could generate its own categories instead of relying on a fixed set as above. To accomplish this, the categorization mechanism described in chapter 7 was used to process the sensory input before it was handled over to the chaining mechanism. Since the left and the right sensory signals were both included in the sensory representation, the creature could potentially recognize both its location and its heading as before. The matching threshold was set to make the creature generate approximately the same number of categories as it had used in the previous simulations.

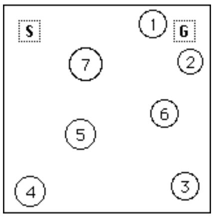


Figure 8.3.9 The environment used for response chaining with dynamically created place categories. Seven stimuli, and the start (S), gives off different smells that are recognized by the creature.

Figure 8.3.9 shows the environment with eight different stimuli (the start, S, is one of these). These will generate a sensory input which is unique for each location in the environment and changes smoothly with the location of the creature. The input to the categorization network consists of the array of sensory signals from both the left and the right sensors, and was normalized as described in section 7.2.

Figure 8.3.10 presents an overview of the current system (compare figure 7.2.4). The sensory input is first categorized by the categorization network presented in chapter 7. New categories are generated when the best category does not fit the input sufficiently well. The categorization network will, thus, generate a Voronoi tessellation of the sensory space, which is subsequently used instead of predefined place categories. In the procedural learning system, the place categories are associated with the appropriate behavior using the network described in section 8.3. The output from the learning system is finally used to activate the correct behavior module.

The connections in the overall system are mainly feed-forward, but there are also connections in the opposite direction. The procedural learning system can request the generation of a new category when the actual reward is less than the expected as described in section 7.2. The matching module shown in figure 7.2.4 is here considered a part of the procedural learning system.

There are also back-connections from the behavior modules to the procedural learning system which represent the performed behavior. It is, thus, assumed that behavioral competition takes place among the behavior modules and not in the learning system itself.

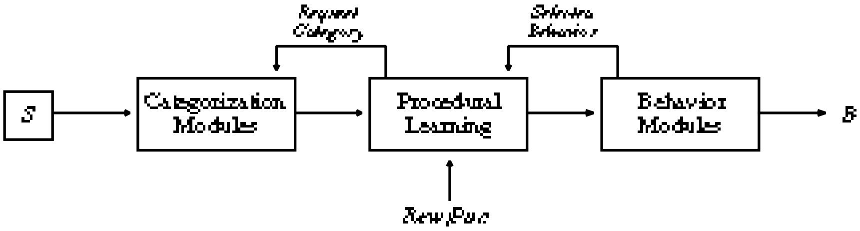


Figure 8.3.10 The architecture of a procedural learning system with categorization. The categorization modules generate catagories using the network shown in figure 7.2.2. The procedural learning system consists of the network in figure 8.3.1. Learning in this subsystem is controlled by reward and punishment. The procedural learning system can request new categories from the categorization network when the actual reward is lower than the expected, as shown in figure 7.2.5. The behavior selection is assumed to take place within the behavior modules and sent to the learning system to control learning.

The behaviors shown in figure 8.3.4 were used also in this simulation. Since new categories were generated as soon as they were needed, the categorization was opaque to the procedural learning system. Since approximately the same number of categories were generated as before, the learning did proceed almost identical to that shown in figure 8.3.6. Within this environment, the back-connections from the procedural learning system to the categorization network did not appear to play any role. Although some categories would be created using these back connections, the speed of learning was not affected if these were removed. We suspect that they will be needed in a more complex environment, however (compare section 7.2). Especially if we allow the creature to generalize (see chapter 9).

To summarize, we have shown how a categorization network in combination with a chaining mechanism can form its own categories which it subsequently uses to learn behavior sequences. This view of learning is consistent with the division into separate systems for categorization and behavioral learning described in section 2.5. Also note that reward is used only within the procedural learning system and not in the categorization network. If we assume that learning within the categorization network is slower than within the procedural system, we would obtain a faster learning process if the creature is first exposed to the environment without being rewarded. This would, thus, be an example of latent learning.

Problems and Limitations

There are a number of problems in the procedural learning presented above. The first can be found in the categorization module. It is necessary that the creature generates sufficiently many categories to give a good representation of where it is, and in what direction it is heading. With only the eight stimuli available in the simulator, it proved to be hard to place the stimuli in such a way that the generated categories would be a good representation of the environment. In many cases, the generated categories would only recognize which stimulus was the closest rather than the whole stimulus configuration. The reason for this is that the intensity of the smells decays very rapidly as the creature moves away from a stimulus. The stimulus closest to the creature has, thus, usually a much larger intensity than the others. As a consequence, this stimulus will dominate the created categories completely.

This problem was solved by placing the stimuli sufficiently close to each other. The drawback of this solution, apart from being a specially prepared environment is that only a fairly small environment could be used. To solve the problem more generally, a better type of place–categories are needed. However, the model could easily be extended with real place–categories. Any of a number of models could be used here. For example the models proposed by Zipser (1985), Prescott and Mayhew (1993), Schmajuk and Blair (1993), and Touretzky and Redish (1995). This problem is, thus, sensory rather than architectural.

The second problem has already been discussed in the end of chapter 5. The chaining mechanism models interstimulus interval effects rather crudely. This demands that the stimulus category changes between each time step, which in turn requires much more categories than would otherwise be necessary. We believe that by including a mechanism for ISI-effects, learning will be much faster (see appendix C). In the simulations described above, the learning rate at the negative side of the reinforcement module was half that of the one at the positive side to compensate for the times when the category would not change as the creature moved. While this solution appears to work fairly well, it is still not very satisfactory.

The final limitation of the procedural learning system is that it cannot generalize at all since it uses a localist representation of location and direction. This means that one node in the network codes for each specific location. It is likely that a better performance could be generated by letting the creature use distributed representation instead. This is an idea we will again return to in chapter 9.

8.4 Expectancy Learning

Procedural learning is very general since it allows almost any behavior to be acquired given sufficient time to try out different behaviors. The drawback is, of course, that it is very slow. In this section, we will investigate how the expectancy network presented in chapter 5 can be used to learn an environment much faster than by procedural learning.

The model of spatial orientation based on expectations we will present is similar in spirit to that of Schmajuk and Thieme (1992). The network used here is different in some respects, however. One important difference is that the present model is based on an expectancy network instead of an ordinary recurrent network. The expectations used are, thus, considered to be established through classical conditioning during exploration of the environment as described in section 7.4.

We will assume that the appetence behavior of the creature is divided into two alternating phases. In the first phase, the creature samples the various stimuli in the environment and calculates how well each of them predicts the desired goal. In the second phase, the creature approaches the stimulus which is the best goal predictor. (Compare the combined orientation and approach system described in section 4.2.)

Figure 8.4.1 presents a small network which uses an expectancy network to calculate how well each stimulus in the environment predicts the desired goal. The purpose of this network is to select the best goal-predictor and make the creature approach it (see also appendix E).

Let us assume the creature finds itself in a situation where two stimuli, CS_1 and CS_2, are present. The creature is assumed to evaluate each stimulus in turn in the following way. First, a brief pulse, ΔCS_1, is sent to the network. This signal will activate the node v_1 which will stay active as long as CS_1 is evaluated. The pulse will also enter the expectancy network where it will be propagated through the network and generate expectations from CS_1. These new expectations will then enter the expectancy network again and generate further expectations from the new nodes, and so on. The expectations which return to the network again will be

Let us again consider the situation described at the beginning of chapter 4 (see figure 4.2.3 and 4.2.4). The creature wants to move from its current location S_S to a goal location, which we will call G, and samples the two possible routes in turn to see which path is shortest.

First it evaluates path X where stimulus S_2 is used as the first subgoal. The stimulus S_2will generate an expectation of the stimulus S_3, which in turn will generate an expectation of S_4, and so on, until the goal representation is activated. Since there are five associative steps from S_2 to the goal, and each expectation is discounted by a factor δ, the goal prediction given by S_2 will be δ^5.

Now the creature evaluates the other possible route, called Y, in the same way. The recurrent expectations which activates the goal representation from stimulus S_{II} will be discounted by δ^2 since two associative steps are needed. Since δ^2 is larger than δ^5, the creature will subsequently chose path Y over path X. It seems appropriate to consider the expectations formed as an internal representation of the life-space of the creature. In chapter 9, we will discuss how this mechanism relates to planning and problem solving. Below, we investigate how our artificial creature can use recurrent expectations to behave in a simple environment.

Spatial Learning with Expectations

Figure 8.4.2a shows the behavior of the creature when it has learned the simple environment in figure 8.3.7. In figure 8.4.2b, the corresponding life-space is shown. The creature will correctly choose to approach the goal directly from all locations in the environment. For this type of learning, it is not meaningful to show how performance improves over subsequent trials. After an initial exploratory phase, the behavior will immediately reach a very good level.

Figure 8.4.2 Behavior based on expectancies. (a) The first trial when the environment is explored. (b) The creature performs very well even on the second trial. The main problem for the creature is not to find the goal, but to avoid walls. A very simple obstacle avoidance module was used in this simulation which generates the inefficient turn around the final corner before the goal.

With learning as fast as this, we can investigate much more complex situations. A second simulation was run in which the creature was placed in the maze presented in figure 2.15.1. A similar maze was used by Tolman and Honzig (1930) to show that the behavior of rats is guided by a cognitive map rather than by habit, and that the animals can show “insight” into the maze when needed. In the simulation, we tried to mimic the experiment by Tolman and Honzig as closely as possible. Since the sensory system of our creature is much more primitive than that of a rat, some differences are necessary however. To let the creature solve the problem, it is necessary that at least one stimulus is placed in each corner of the environment as shown in figure 8.4.3a.

In the first phase of the experiment, the creature was allowed to explore the environment as shown in figure 8.4.3b. During its exploratory behavior, the creature will set up expectation about which stimuli are close to each other by classical conditioning (see appendix I). After this phase, the creature is forced to select stimulus 1 as a goal and is tested under one of three conditions.

Under the first condition, the creature is simply placed in the start box in the bottom of the maze. As can be expected, the creature starts to approach stimulus 1 by the direct path (figure 8.4.3c). Since the creature was initially directed away from the goal, it first had to turn which is the reason for the small deviation from the shortest path shown in the simulation.

In the second condition, the shortest path is blocked between stimuli 7 and 8 as shown in figure 8.4.3d. As a result, the creature will choose the second best path to the goal instead, that is 7–6–8–2–1. Initially, the creature chooses stimulus 7 as the best goal predictor, but while it attempts to approach it, the expectation from stimulus 7 to stimulus 1 will extinguish. At this time, stimulus 6 will become the best goal predictor and the creature will turn toward it instead and choose the path to the right in the maze. As described in chapter 2, this is also the behavior one would expect from a reinforcement learning mechanism.

Under the final condition, the path from start to goal was blocked as shown in figure 8.4.3e and f. As before, the creature starts to approach the goal by the shortest path until stimulus 8 loses its role as best goal predictor. At this time, the creature senses two other stimuli, 6 and 7. Depending on which of these stimuli are the closest, the creature will either choose the behavior in figure 8.4.3e or that in f.

According to Tolman and Honzig (1930), this is an example of insight on behalf of the creature. If it had used a stimulus–response chaining mechanism like the one described earlier in this chapter, the creature would not know what to do when the path was blocked and when it had reached stimulus 7 again after the behavior from stimulus 8 had become extinguished, it would not have chosen the second best path as under the second condition. It would, thus, have tried the path 7–6–8–2–1 even though it had experienced the blocking between stimuli 8 and 2 already.

The creature could only choose the longest path immediately if it had knowledge about the layout of the environment, and not only about what behavior to perform where. With such knowledge, it could infer that the blocking on the shortest path would also be a blocking on the second best path.

This learning mechanism is, thus, a large step from the stimulus–response chaining described above. Also note that for expectancy learning, it is much more reasonable to use stimulus–approach behavior than fixed responses. In the network presented here, this made it possible to form expectations between stimuli without any regard for how the approach behaviors would be executed.

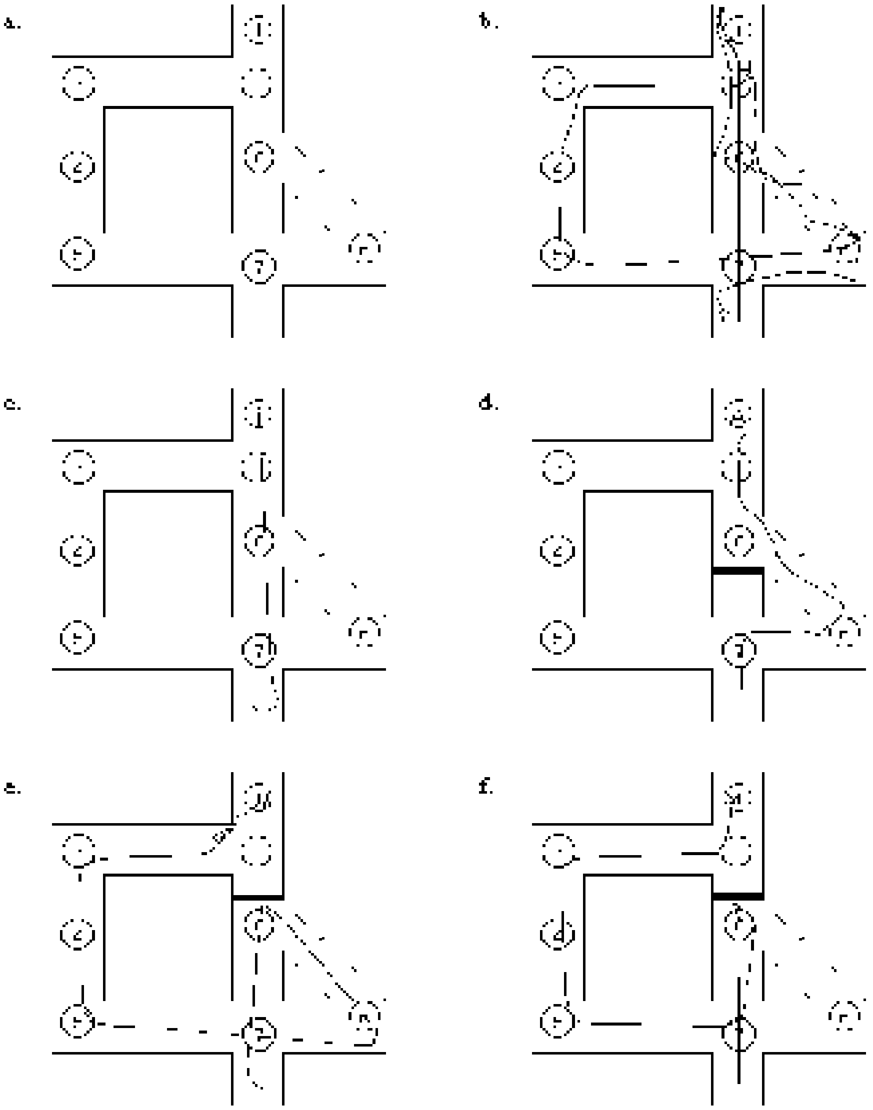


Figure 8.4.3 A simulation of the experiment run by Tolman and Honzig in 1930. (a) The location of the different stimuli. (b) The exploratory phase. (c) Behavior when

all paths are free. The creature chooses the shortest path from start to goal (1). (d) The creature selects the second best path when the shortest is blocked. (e) One of two behaviors generated when the shortest and the second shortest path are blocked at the same place. When the creature finds the shortest path is blocked, it will select the second best goal predictor instead. In this simulation, stimuli 6 and 7 are equally good goal predictors. As a consequence, the creature will sometimes choose the behavior in figure (e) and sometimes the one in figure (f).

In the spatial domain, expectancy learning is obviously more efficient than stimulus–response learning, but this type of learning also has its share of problems. These are mainly a consequence of limitations of the expectancy network. In the simulation described above, the creature would alternate between two paths from start to goal when the direct route had been blocked (figure 8.4.3e and f). In this simulation, the path 8–7–5… is always shorter than the path 8–6–7–5…, but the creature would very often choose the longer path instead. The reason for this is that the expectancy network does not convert physical distance into psychological distance in the optimal way. Since stimuli 6 and 7 have both been conditioned to stimulus 5, they are both equally good goal predictors allthough the distance from 6 to 5 is longer than the distance from 7 to 5 (see appendix E). To handle this problem, physical distance must somehow be included in the expectations.

Including such information is not very hard if the absolute smell intensity of each stimulus is known, but it is hard to see how such information could be generated by other modalities. What is needed here is a representation of the distances between the different stimuli which is independent of the sensory modality used. Such distance information could possibly be generated by dead reckoning as the creature moves from one location to the next during the exploration phase or by including systems for distance calculation within each modality. In either case, the association formed in the expectancy learning should be a function of the perceived distance from one stimulus to the next (compare section 5.10).

8.5 Conclusion

In this chapter, we have considered two radically different ways to learn a behavior sequence in the spatial domain: procedural and expectancy learning. It was shown that expectancy learning is much faster and can also be used to solve simple problems. Procedural learning, on the other and, is slow and generates rigid behavior that cannot be easily changed.

In view of these properties of the two systems, one may ask whether we have bothered to present the procedural learning system at all. There are a number of reasons for this. The first is that it is clear that a system of this kind exists in real animals (Shimamura 1990). When a task has been tried on a large number of occasions, animals will typically start to behave as if a stimulus–response mechanism is involved. This is usually called *overtraining* (Gallistel 1990). One possible role of the procedural learning system in this case may be to relieve the expectancy network of controlling repetitious behaviors, and to free it for more important tasks (see chapter 9). We may also view the expectancy system as a tutor for the procedural learning system. When the internal incentives generated by the procedural learning system becomes higher than those of the expectancy system, the control of behavior will automatically shift to the procedural system. The relation between these two systems will be further discussed in chapter 10.

Another possible use of the procedural learning system is to modulate the behavior generated by the more cognitive system. The expectancy network generates a fairly good behavior very quickly, but it is not optimal. By combining both types of systems, learning becomes fast while still being able to reach an optimal level eventually. In this case, the expectancy network can be seen as a search-heuristic for the procedural learning system. Since the expectancy system only executes approximately correct behaviors, the procedural learning system will not have to train on a large set of useless behavior sequences as when it is used on its own.

It is interesting to note that the network for expectancy learning is not much more complex than the one for procedural learning. The chaining of stimulus–response associations is in no way simpler than the learning of expectations. In the models presented here, the same type of reinforcement module was used as the main building block for both systems.

The present chapter can, thus, bee seen as a contribution to the classical controversy between the two views of learning. In the tradition started by Thorndike, learning is seen as the acquisition of habits much like the behavior chaining described above. In the other tradition, which is usually associated with Tolman (1932), animals are assumed to acquire knowledge about their environment rather than habits. If learning was, in fact, the acquisition of knowledge, the behaviors of animals in various experiment appeared much easier to explain than with a habit theory. There was one large problem with the cognitive theory, however. It seemed impossible to give a mechanistic account for how the knowledge was converted into behavior.

It took many years before a mechanistic model was presented which could explain the purposive behavior of animals (see Gallistel 1980). The first model was probably presented by Deutsch (1960), and used a mechanism which is in many respects similar to the one described above. This model was also different from the one presented here since associations would not flow from potential subgoals to the goal, but instead from the goal to the current location of the creature. As already mentioned, the model presented by Schmajuk and Thieme (1992) is more similar to the expectancy net used here since it too selects a number of potential goal predictors and evaluates them sequentially.

We can compare the two systems presented here with the distinction between procedural and declarative memory (Shimamura 1990, Squire 1992), although these systems are naturally much more complex in real animals. The view promoted by the present chapter is, thus, that both these theories are valid, but for different types of learning. We also believe that the use of expectations rather than simple associations will make it easier to understand how planning and problem solving abilities can be seen as a more advanced form of recurrent expectations. This is something we will return to in the next chapter.

Chapter 9

Cognition

9.1 Introduction

The mechanisms we have constructed in the previous chapters have all been directed toward fairly simple problems in the spatial domain, and it is now time to ask what relevance these systems have for higher cognitive abilities. It is our belief that many cognitive abilities share the structure of the tasks we have so far discussed, and in this chapter, we will outline a number of extensions that can be made to the architecture.

We will argue that the systems developed so far constitutes a solid foundation for a full-fledged cognitive system. We will point out a number of problems with the architecture and sketch how the perceptual system can be extended to handle more modalities and representations at various levels of generalization.

The expectancy network introduced in chapter 5 can be used as the starting point for the internal environment which is the corner-stone of planning and problem solving. When internal categories are formed for recurrent expectations, chunkin becomes possible, which in turn allows planning at different levels of specificity. All these extensions require that the creature uses multi-modal representations however, and this will be the first topic of this chapter.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

9.2 Multi-Modal Representations

The creature introduced in chapter 3 has very limited sensory abilities. Everything it knows about the world comes from 28 simple sensors. While this is a fair amount compared to the number of sensors in many robots, it is very low compared to almost any real animal. The amount of information extracted from the sensors is also very limited. There are, thus, a number of possible extensions that can be made to the model. Below, we will consider some of these to make the discussion of cognition more general.

Olfaction

The modality we have called sense of smell is a very pale copy of the olfactory receptors used by real animals. An extension of the model would use many more receptors. In a more realistic model of smell diffusion, it would no longer be possible to use the difference between the signals from the left and right sensors to guide locomotion directly. There are usually no static smell gradients in the environment to be picked up directly as we have so far assumed. Much more processing is necessary before the direction to the source of the smell can be established. This may be the reason why most animals do not have their olfactory sensors very far apart. There exists plenty of counterexamples of course. One is the ants who use pheromone trails for homing. Since pheromone trails are much more stable than substances diffusing through air, it is more favorable to use two sets of receptors in this case.

This view of olfaction is easily reconciled with the model of optimal approach behavior presented in chapter 4. Such a behavior consists of the two phases orientation and approach. To use realistic smell signals, an animal would have to divide its behavior into successive phases of the olfactory based orientation toward the goal, and the approach behavior. The approach behavior could possibly be blind in this case, that is, it could consist of locomotion straight ahead. Alternatively, it could be controlled by dead–reckoning (see Gallistel 1990).

Another point to mention about real olfactory sensors is that they have varying degree of specificity (Shepherd 1991). Some sensors react to very few chemical substances while others fire for a much broader range of inputs. As we have already seen in relation to perceptual categorization in chapter 7, nodes with larger specificity should have a larger influence on later processing stages than nodes with lesser specificity. In section 9.3 below, this will be formulated as a general principle.

The Somatic Senses

Our creature has four sensors that can be called somatic. The first two are the whiskers at the left and the right of the body. The third is the collision detector and the fourth is the single pain receptor. The two whiskers show an important principle of the somatic sensors of real animals: They are spatially distributed on the body. Using only two whiskers, the resolution is fairly limited, but it has already been seen how the whiskers can be used for obstacle avoidance in a fairly efficient manner, since the sensors produce spatial information.

The collision sensor was included to handle cases where the whiskers do not react, but this is not a particularly good solution. Since the collision sensor does not give any directional information, the best the creature can do when it reacts is to turn at random. A much better system would use many more whiskers or other tactile sensors everywhere on the body. Since the spatial pattern of activation would inform the creature of where an obstacle is and how large it is, a much more efficient avoidance behavior could be constructed. In some animals, such as the snake, it is even possible to let the avoidance behavior be calculated locally in the spinal cord using lateral inhibition (Hirose 1993).

The representation used in the tactile sensors can be considered as a *representation by place* (Martin 1991). This means that a node has the same meaning each time it fires. The signal from the whisker on the left means that there is something touching on the left. The intensity of the signal directly codes for the intensity of the touch and the signal always has the same meaning. There is no message coded in the signal other than intensity, which means that no decoding is necessary.

A second property of the somatic representation is that it is *spatiotopic*, that is, arranged according to the spatial location of the sensors. Sensors that are close to each other on the body are represented by nodes that are physically close in the brain (Kandel and Jessel 1991). We used this property in the construction of many of the reactive behaviors described in chapter 4. Again, the resolution is very low since the creature can only distinguish between left and right. A better spatial resolution would make many tasks much simpler. This property will be considered again below in relation to orienting behavior and multisensory mappings.

In most animals, the spatiotopic representation does not directly tell the animal where in space an object is located since this also depends on how the different parts of its body are positioned at a certain time. The representation is in local body coordinates and it appears sensible that this representation should somehow be converted into a global egocentric coordinate system in many cases.

Vision

To make the model simple, we have deliberately avoided vision which is otherwise the most powerful sensory system used by real animals. Visual input has a number of interesting properties that would be useful within the model presented above, however.

The first is that it works at almost any range. It can be used to recognize an obstacle long before the creature comes into contact with it. The same is true about the identification of goals, landmarks and other cues. The inclusion of vision makes all these processes simpler and more effective once the visual image has been analysed. The analysis of the visual image itself is however a far from trivial problem, and this is why this modality has not been considered above.

The second property of vision which is of importance in this context is that it is spatiotopic in a way similar to the somatic representations. For vision, however, the representation is arranged according to the location on the retina and not the body. This type of representation is usually called *retinotopic* (Martin 1991). Like the somatic senses, the coordinate system used in the visual representation is a function of the orientation of various parts of the body, most notably the position of the head and the eyes. To calculate the location of a visual cue in egocentric, or even allocentric, coordinates, some form of coordinate transformation is necessary. Andersen and Zipser (1990) have proposed a computational model of how this process may be implemented in the posterior parietal cortex.

It appears that the mammalian brain solves this problem by dissociating the calculation of spatial location from the identification of a visual cue (Mishkin, Ungerleider and Macko 1983). By centering the cue on the fovea, the brain does not generally require that the representation is translation invariant. As a consequence, the calculation of spatial location can be made without reference to the actual cue. Only the vergence angles between the eyes and the position of the head need to enter the calculation. This makes it much simpler to use a spatiotopic representation for the location of a cue.

Orientation movements can, thus, be made with respect to the representation of the location of a cue without involving the representation of the cue in itself. That is, orientation movements will automatically be generalized to all visual cues. It seems that some similar mechanism must operate for the somato-sensory modalities, but we are not aware of any empirical evidence that either supports or contradicts this idea.

The third property of the visual system which fits nicely with the model presented above is its organization as parallel visual control systems (Goldberg, Eggers and Gouras 1991). The visual system is equipped with behavior modules, possibly involving the whole body, which perform some specific visual behavior such as a saccade movement, smooth pursuit or a visual orienting reaction. In smooth pursuit, the eyes track a moving object to keep it fixed on the fovea. This is, thus, a form of approach movement as it was defined in chapter 4. The visual orienting reaction is, of course, just another kind of orienting reaction that also fits well within the framework presented in chapter 4. Like the behavior modules presented above, the visual behaviors can be used by different engagement modules for varying purposes.

Multi-Modal Interaction

The reason for introducing many modalities is that cognition is often concerned with many types of sensory information at the same time. Such representations are usually called *multi-modal* to indicate that they are made up of many modalities (Stein and Meredith 1993). In other cases they are called *supra-modal* to show that they are, in a sense, above a single modality (Murray 1990). According to Sjölander (1993), the emergence of intermodal representation is one of the greatest break-throughs in the evolution of cognition and consciousness.

The important property of multi-modal representations is that information from one modality can be transformed into another. By touching an object, we can imagine how it looks, and by looking at it, we can imagine how it would feel to touch, and even what sound it would make when we touch it or if we would drop it on the floor. How this sort of transformation is performed is not very well known except in the case of orienting movements, so we will concentrate on this type of multi-modal interaction.

In the case of the orienting reaction, the different modalities interact through the use of a common spatiotopic code (Stein and Meredith 1993). Since it is only the location of an object that is important for an orienting movement, all modalities can easily be mapped onto this common spatial code (see Balkenius 1995). It does not matter whether an input comes from the visual, auditory or somato-sensory system since each of these modalities contains information about the location of a stimulus, and it is only the location that needs to be represented in the orienting system.

The main problem solved by the orienting system[1] is to choose which stimulus to attend to when many are simultaneously present (Johnston and Dark 1986). This could be done by competition using the choice mechanisms presented in chapter 3. In this choice, the intensity of a stimulus together with a count of the number of modalities involved determine which stimulus will be chosen. The choice can potentially be biased by expectations about the current situations generated by an expectancy network as described in chapters 5 and 8. One role of the expectancy network is, thus, to habituate the orienting reaction to stimuli that have already been attended to (see also section 7.4).

Submodalities and Perceptual Dimensions

While multi-modal interaction is important, one should not forget the interaction which takes place within a single modality. This interaction occurs between different *submodalities*. Here, we will use vision as an exanple. It is well known that the analysis of the visual input is segregated into a number of submodalities, the most

[1] We are referring here to the system governing the orienting reaction and not a system for spatial orientation.

important being motion, form and color (Livingstone and Hubel 1988, DeYoe and Van Essen 1988). Since all submodalities exist in parallel, a visual input will give rise to representations in all submodalities at the same time. This means that we cannot rely on a specific node for each stimulus in the environment. If a stimulus is analysed in three submodalities, the minimal number of nodes used for the output is also three, but it can obviously be much larger. As a consequence, the perceptual system must always deal with configurational representations. A single stimulus will never be mapped onto a single node, but onto a whole set. This is usually called a *distributed* representation (Hinton, McClelland and Rumelhart 1986).

Another type of submodality that is required in vision is the representation of stimuli at different scales (Witkin 1983). This kind of submodality also exists in the olfactory system (Shepherd 1991). The result of these submodalities is that every stimulus will be represented at different levels of specificity even before any categorization takes place. A node with high specificity will react to very few stimuli while nodes with lower specificity will react to a larger set of stimuli. It is possible to consider each submodality as a dimension along which every stimulus is categorized. We will call the representation in all submodalities taken together the *sensory schema*. This is shown in figure 9.2.1.

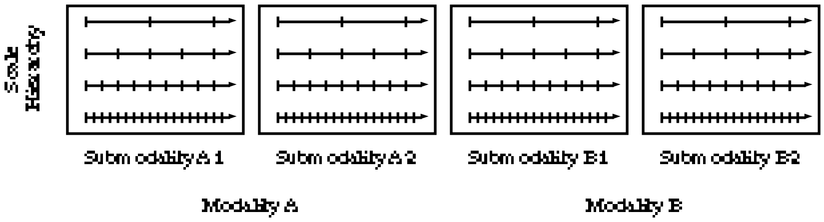


Figure 9.2.1 The distributed sensory schema consists of the collection of representations in all modalities. Each modality is divided into a number of dimensions or submodalities that may represents the stimulus properties in different scale-spaces or levels of specificity.

The sensory schema is a collection of signals that are all active when the sensors of the creature are directed toward a specific stimulus. To discuss such schema we need to distinguish between stimuli external to the creature, and the sensory schema they activate, which is internal. Sometimes, these two aspects of a stimulus are referred to as the *distal* and *proximal* stimulus respectively (Gray 1975).

The conclusion that can be drawn from this section is that it is necessary to consider all stimuli as configurations of the representations in all the individual modalities. It is, thus, too simplified to consider the signals entering the cognitive system as representations of individual stimuli as we have done in the previous chapters. The representation of a conditioned stimulus, CS_i, must be replaced with the sensory schema α_i it activates in all the sensory modalities. This will make the mechanisms for categorization and association much richer as well as allowing for generalization (see Balkenius 1994c).

9.3 Categorization

In this section, we will discuss how the categorization mechanisms presented in chapter 7 can be extended to handle sensory schemata instead of stimulus cues. The perceptual mechanisms presented in chapter 7 assumed two levels of representation. The first was the sensory level, at which the output level of a node has a monotonous relation to the response level of a single sensor. The second level of representation consisted of configurational categories. These are essentially collections of sensory signals that have a special meaning. Representation of the first type will be called *first–order categories*, while the second type of representations will be called *second–order categories*. How can the categorization mechanisms described above be extended to handle sensory schemata?

In chapter 5, we introduced the problem of negative patterning (see figure 5.5.3). In this situation, the conjunction of two sensory cues, say, S_0 and S_1, should produce different predictions of the environment than each of the cues taken together. This requires that a configurational category is created for the combined stimulus, $S_0{\bullet}S_1$. It is also necessary that the expectations from the configurational category can compensate for those from the individual stimuli. As we saw in chapter 7, expectancy learning could automatically generate the required associations.

Making the transition from stimulus cues to sensory schemata requires that all stimuli are handled as configurational. Given a sensory system with two modalities A and B, a single stimulus, S, will activate one or more representations in each modality. Let the representation in modality A be called α, and the representation in modality B be called β. The representation of S will then be the conjunction of the representations in the two modalities, that is, $\alpha{\bullet}\beta$. This representation must obviously be treated in the same way as the configurational stimuli were treated above. The initial predictions made from the stimulus representation will handle each modality independently until expectations are not met. In this case, a new category will be created which will represent the sensory schema $\alpha{\bullet}\beta$ (see the simulation in figure 7.2.5). This second–order category will be denoted by $\chi_{\alpha\bullet\beta}$.

Let us consider a concrete example where stimulus S_0 consists of a blue circle. We will assume that this stimulus is analyzed in the two visual submodalities for shape and color. The signals from the shape system will represent that the stimulus is round and the signals from the color system will represent the color blue. Let us further assume that the presentations of blue shapes or round shapes have always been followed by a high pitch tone, called S_1.

When presenting a round shape, the shape representation will cause the expectancy system to predict that the high pitch tone, S_1, will follow. This prediction will also be made independently from the color representation, when a blue shape is presented. If the expectation is not fulfilled, say, if the presentation of the blue circle is followed instead by a low tone, called S_2, a mismatch occurs. The predictions made from the sensory schema BLUE•ROUND was incorrect. This will cause a new category to be created that lets a specific node react every time a round blue shape is presented (see section 7.2). This category, that we may call $\chi_{\text{BLUE•ROUND}}$, will subsequently predict both the non-occurence of the high pitch tone, S_1, and the presentation of the low pitch tone, S_2, as shown in figure 9.3.1.

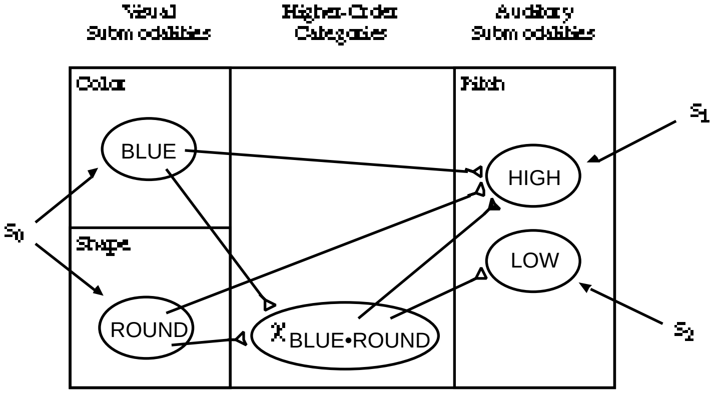


Figure 9.3.1 Category formation as the result of negative patterning between different submodalities. The presentation of a blue circle predicts a low pitch tone, S_2, although all other shapes that are either round or blue predict a high pitch tone, S_1. This will cause a second–order category to be created that represents this specific exception.

Note that the second–order category is more specific than the nodes coding for the individual submodalities. The level of specificity will increase with the number of submodalities that must occur together for the category to match. With more specific categories, the predictions made can also be more specific. This is, thus, an example of a learned scale hierarchy. Also remember that we are still talking about neural activity patterns although the example is beginning to look like an attribute–value structure.

This is an example of category formation by specialization (Charniak and McDermott 1985). Another possible type of category formation is categorization by generalization. In this case, a set of specific categories learn the same expectations. Two stimuli, S_0 and S_1, activate the perceptual schemata α and β respectively. Assume that each of these schemata is independently associated with a category node, γ, which already exists in the system. Since both α and β can activate γ on its own, it seems appropriate to consider this category as a generalization of α and β.

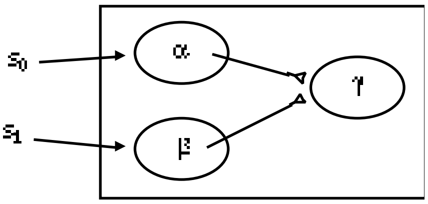


Figure 9.3.2 Categorization by generalization is handled by an associative process. No new category need to be created in this case.

A fairly simple extension of the mechanisms we have described earlier in this book would, thus, handle categorization based on both specialization and generalization. New categories are only created during specialization which is natural since generalization assimilates the new instance to a category that already exists.

9.4 The Internal Environment

The expectancy network introduced in chapter 5 can be used as a starting point for what might be called an internal environment. This environment is used to try out actions internally before they are confronted with external reality (Gulz 1991). As we will argue below in section 9.5, it can be used both for planning and problem solving. In this section, we discuss how the expectancy network would have to be modified to serve as an internal environment.

Temporal Expectations

The temporal expectations considered above all require a fixed time delay between the cue and the predicted event. A prediction of this type will be written as,

$$S_i(t) \to S_j(t+1). \qquad (9.4.4)$$

Using this type of representation, predictions over longer periods of time could only be made using avalanches started by a sensory cue or by secondary conditioning (see chapter 5). In a more general system, predictions must be made over arbitrary periods of time. It is not sufficient to represent that the presentation of S_i at time t will predict the presentation of S_j at time t+1. A more useful expectation would represent, not only that S_j will follow S_i, but also when. Representations of this type would look the following way,

$$S_i(t) \to S_j(t+\Delta_{ij}). \qquad (9.4.5)$$

Here, Δ_{ij} is a parameter that is specific for each pair of stimuli. A further extension would replace the single parameter Δ_{ij} with a function of time that assigns a probability to the statement $S_i(t) \to S_j(t+\delta)$ for each value on δ. This function can be represented either directly, or as a set of parameters.

One way to parametrize the representation is to represent the prediction as an interval $[\Delta_{ij}-\xi_{ij}, \Delta_{ij}+\xi_{ij}]$ around a mean value represented by Δ_{ij}. This requires, of course, that S_0 has preceded S_1 at least once. If it has not, the values of Δ_{ij} and ξ_{ij} have no meaning, and this suggests that we should also represent the certainty of both Δ_{ij} and ξ_{ij}. This certainty, c_{ij}, should be zero initially, and increase as the estimates of Δ_{ij} and ξ_{ij} improve. This is not sufficient, however.

To know whether S_i really predicts S_j, we need to weigh it against the probability that S_j is present without being preceded by S_i. If, for instance, S_j is always present, S_i has obviously no predictive power and there should be no association between S_i and S_j. A similar effect occurs in latent inhibition experiments where the presentation of a CS on its own retards the subsequent development of an association with an US (Mackintosh 1983). In this case, the pre-exposure to the CS makes the probability that the US will follow smaller than if the CS had never been presented before. This suggests that latent inhibition can be explained with a mechanism that estimates the probability of occurance of each stimulus and compares it with the probability of each CS–US pair (Gallistel 1990).

Balkenius (1992) has presented a neural network architecture that is able to extract the amount of information one node transmits about the state of another. Although it was not recognized in that paper, the proposed network model solves the problem of latent inhibition in the case of simultaneous conditioning. However, the model is not easily extended to temporal predictions, and this is what is required for a useful internal environment.

An extended creature will, thus, need to form predictions that are based on transferred information and not on temporal contiguity. These predictions must code both *when* a stimulus is likely to follow another and with what *certainty* this prediction is made. As in the case of categorization, these predictions must also be made for distributed schemata and not for local stimulus representations, since it is necessary to represent predictions from one submodality to another.

Working Memory

The segregation of the sensory analysis into ‘what’ and ‘where’ systems (Mishkin, Ungerleider and Macko 1983) has some interesting consequences. Let us assume that the creature is equipped with a content addressable memory that works in the following way (see section 3.3). When an orienting reaction is made toward an unexpected stimulus, the sensory schemata represented in the ‘what’ and the ‘where’ systems are associated together in the content addressable memory.

This memory can be seen as a generalization of the working memory for novel foods introduced in chapter 5. In that simple working memory, only the sickness sensor could recall the stored memory. For a more general working memory, any sufficiently specific subschema from any modality should recall the entire sensory schema with which it was stored. For example, the activation of a place representation will read out the expectations of the stimulus that will be perceived at that location. In the same way, the activation of some other sensory property will read out the location at which that property can be found (figure 9.4.1).

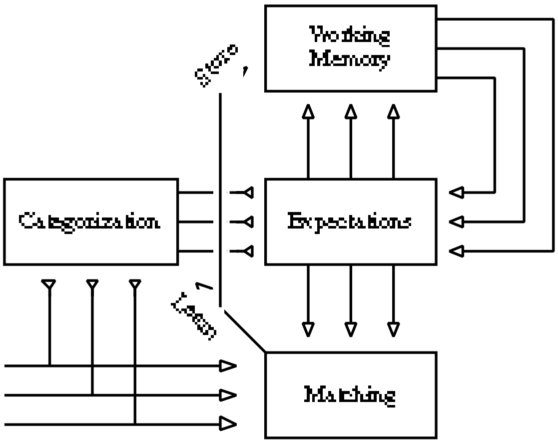


Figure 9.4.1 The expectancy network equipped with a temporary working memory for binding of sensory representations. The general expectations in the expectancy network are enhanced with specific predictions in the working memory.

Content addressable memories have been much studied within the neural network literature and there are many models to choose from (see, for example, Hinton and Andersson 1981, Kohonen 1984, and section 3.3). In many respects, the working memory has many of the properties of the expectancy network presented in chapter 5. One may conceive of the working memories as a collection of expectations formed by simultaneous conditioning. The problems associated with negative patterning will also show up in the working memory, which means that it too will require a categorization mechanism. These problems are so severe here that it would be possible to form categories of all new schemata even without checking whether it is required or not (Rolls 1990). Since it is required that working memories are very specific, high-order categories will always be needed.

In fact, if the working memory, categorizes each novel stimuli into a unique category in this way, it could also potentially generate a learning signal that could start the creation of higher–order categories in much the same way as the recruitment mechanism in chapter 7.

There are also a number of differences compared to the expectancy network. Since the working memory stores individual stimuli rather that general contingencies, it is appropriate for this type of conditioning to be based on temporal contiguity rather than transferred information. There is also no need for an extinction mechanism. Like in the case of smell aversion, the memories could fade away gradually until they are extinguished on their own accord.

The inclusion of a working memory may potentially solve the so called *binding problem* (Hinton, McClelland and Rumelhart 1986). It can be stated as follows. If sensory processing is segregated into finer and finer submodalities represented at different locations, how does the brain know which sensory properties belong to the same stimulus? The solution we suggest is to bind sensory features together, by associating them with the spatial location at which the stimulus occurs, and store the combined schema in the working memory. In support of this idea, it has been shown that location is a more reliable cue for recall from memory than, for example, color or shape (Johnston and Dark 1986, see also Mishkin, Ungerleider and Macko 1983).

The expectations stored in this working memory must be treated as expectations both of the present and the future. If stimulus S is present at location L when the memory is stored, the working memory must enforce the expectation that this is still the case as long as the memory lasts. When this type of association has been generalized within the expectancy network, it could form the basis for the idea of object permanence (Flavell 1985). An object is a perceptual schema which has the property that it predicts itself to be present in the future. This corresponds to the expectation that a collection of some certain stimulus properties will be present at some external location in the future.

In the radial maze discussed in section 2.8, something like the kind of working memory sketched here would be required. Simply including a working memory does not solve the problem, however. There are still a number of questions that must be answered. For example, it is not clear how the animal knows how to use its memory in the way required by the task. Nor is it apparent how it recognizes what task it is supposed to perform in the first place.

Consequences of Actions

If yet another input is introduced to the expectancy net, it can be used to predict the consequences of performed actions. This new input is the efference copy of the signals sent to the behavior modules or the motor system (see section 2.4). These signals will tell the expectancy network what the creature is doing and will let it take its own actions into account when it predicts the next sensory state. Such predictions are obviously useful since they will be more accurate than predictions that are based only on external stimulation. Another important property of this type of predictions is that they can be used for planning as we will see in the next section.

The recurrent expectations used in chapter 8 to select the shortest path to a goal can be seen as a slightly simpler version of this system. Since we assumed that the creature would only use an approach strategy to reach the goal, it was possible to ignore the consequences of actions. The only consequence of importance to an approach behavior is that a stimulus becomes closer than it was before. It was, thus, possible to ‘plan’ using only sensory information. With the introduction of the type of expectations described here, it will be possible to use all types of behaviors in look–ahead choice. The recurrent expectations described in chapter 8 were of the S–S' kind. With the addition of the efference copies, these associations are extended to the more general S–R–S' type (see sections 2.7 and 2.15).

Note that the efference copies used to predict the consequences of actions are very similar to the recurrent expectations discussed in chapter 8. It may even be possible that the same pathway can handle both types of signals. Recall that the same expectancy network can be used both to produce conditioned behavior and for sensory expectations. By introducing scale–hierarchies also in time (Albus 1991), planning at different levels of specificity is also made possible (see also Newell 1990).

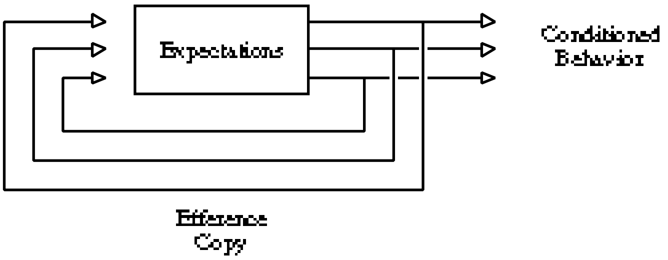


Figure 9.4.2 Predicting consequences of actions by associating efference copies of motor commands with subsequent sensory schemata.

Chunking

Another way to obtain better predictions of the future is to construct categories at even higher levels than so far discussed. The motor behind this mechanism are the recurrent expectations produced by the expectancy network. If we let these recurrent signals enter into new categories in the same way as sensory signals, it will be possible for the system to create categories of categories. Depending on the number of times the expectations have passed through the recurrent connections from the expectancy network to the categorization module, we may talk about categories of the *nth* order. Categories of categories will be of third order. When these are recategorized, the resulting categories will be of fourth order and so on. The recurrent expectations allow categories to be formed of any order. Such categories will, however, only be created when expectations are not met. This prevents categories of higher order than necessary to be formed.

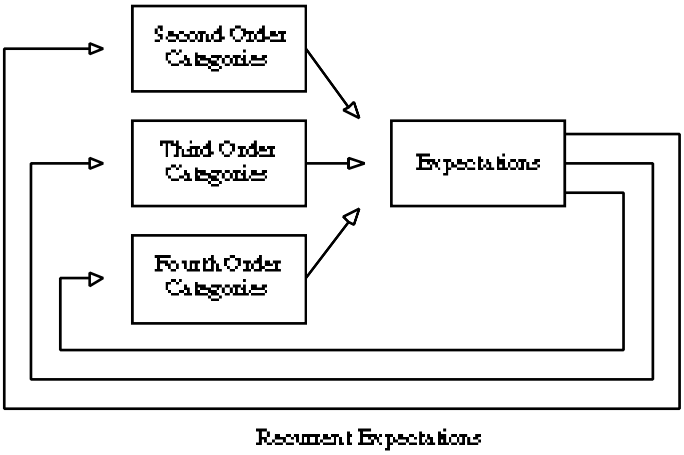


Figure 9.4.3 Chunking by recurrent recategorization

Once the higher–order categories have been formed, they will enter into simple associations within the expectancy network as if they were of second order. Thus, it is not necessary for a sensory schema to be propagated through the recurrent connection *n* times to activate an already established *nth–order* category. One iteration will always suffice (figure 9.4.3).

We suggest that a recurrent recategorization of this type may lie behind the process that is commonly referred to as *chunking* (Newell 1990). If the expectancy network can construct temporal prediction of the kind discussed above, temporal as well as spatial chunking can be handled by the same mechanism. The idea behind chunking is, thus, very simple. Getting the dynamics right is however a far from trivial task.

An alternative to this architecture is to use a fixed number of categorization modules connected in series. This is the approach taken, for example, in the multilayer perceptron (Rumelhart, Hinton and Williams 1986). Since the order of categorization is fixed, the dynamics is much simpler than for recurrent recategorization. Its abilities are more restricted, however.

When a creature has the ability to form accurate temporal predictions both from its sensory inputs and from the actions it performs, it will have the basis for an *internal environment*. It can potentially shut off its sensors and effectors and behave in this internal environment instead of externally. With the introduction of a working memory, the creature will be able to change this environment in much the same way as it can change the external environment. Using the chunking ability, the internal environment can also be handled at different levels of specificity. In the next section, we will describe how the internal environment can be used for planning and problem solving.

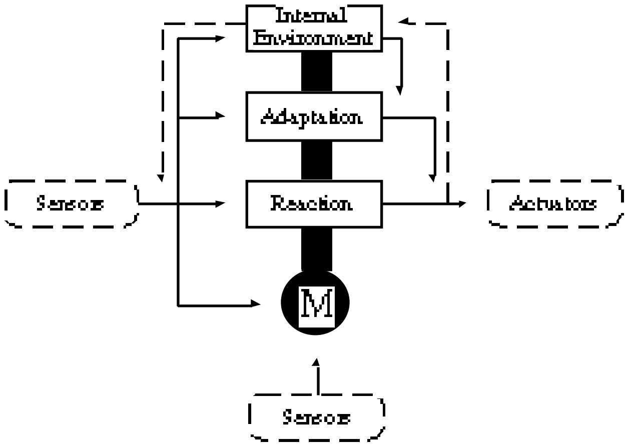


Figure 9.5.1 A functional hierarchy.

9.5 Planning and Problem Solving

If the creature shuts off the output to the motor system and uses the efference copies on their own, it can predict what will happen if it perfoms a certain action. Once this prediction is made, a new action can be tested which is based on the predicted sensory state. In this way, whole sequences of actions can be executed internally.

Figure 9.5.1 shows an agent architecture proposed by Balkenius (1993, see also Gärdenfors and Balkenius 1993). In this architecture, the three functional levels described in section 3.2 are thought of as distinct subsystems that are coordinated by a central motivational system. The lowest layer controls the fundamental reactive behavior. This system equips the agent with a set of elementary abilities that are used as a basis for more complex behaviors. An intermediate adaptation layer controls behavior based on expected rewards. This level is based on reinforcement learning which generates internal incentives that can override the external signals to the reactive level. The top layer learns about the consequences of various actions in the environment. This knowledge can later be used as the internal environment where actions can be tested before they are performed externally.

The upper two layers are similar to the DYNA architecture proposed by Sutton (1992, see also Peng and Williams 1993) and can be used for planning and problem solving. In such processes, the top layer plays a role similar to the world model in traditional AI systems. However, this model is never essential for the behavior of the agent. If this layer is removed, the agent can still learn any behavior in the world given sufficient training. The important aspect of this level is that its representations of the world are *detached* (Gärdenfors 1995), that is, they are not confined to here and now, but can be used independently of the current perceptual situation.

By depending on the already existing systems, a planning ability is the result of adding a single functional layer (see figure 3.2.8). We have an example of a true system property that cannot be localized in one specific module of the agent. With the model system, the agent can be engaged in planning, but this module does not plan in itself. The result of the planning process is to adapt the reactive system to the internal environment, rather than to produce a fixed program to execute (compare Lyons, Hendriks and Mehta 1991).

The architecture presented here is in many respects similar to the one presented by Gulz (1991). Planning is considered as internal simulation of external behavior. Instead of using the external world to generate new sensory information, actions are performed in an internal model. The different actions are simulated in this model, and a new sensory input is generated internally. In this view, the internal model is not independent of the agent itself. It is a module which can generate the sensory information which would be the result if the agent had performed the corresponding action externally.

The internal model is used very differently compared to the models used within the symbol processing paradigm. The model is not a description of the world. As far as our agent is concerned, the internal model is a world. However, to be successful, this internal world must be parallel to the external world. Planning can be seen as behavior in this internal world instead of the external world (Balkenius 1993, see also Gärdenfors 1992, and Gärdenfors and Balkenius 1993). The only difference between a search in the internal world and actions performed externallyd is that all actions can be undone as well as taking much less time to perform. The plan constructed from behavior in the internal world is in no way different from the paths learned from externally tested action sequences. The same mechanisms are used, and the plan is represented and executed in the same way as a behavior that has been practiced in the external world.

The plans constructed using the internal environment are also very different from classical plans. A useful view of the type of plans constructed by this architecture are as resources and not as programs (Payton 1990). Horswill and Stein (1995) have suggested that robot control systems can be classified according to their *commitment* to generated plans. In this scheme, classical AI systems have a high degree of commitment. When the plan has been constructed, it is executed from the first step to the last. On the other hand, the plans constructed by the architecture described here have a minimal degree of commitment, since plans are used reactively. It may even be questionable to call the changes to the reinforcement system plans since they are not explicitly represented as a such.

When the abilities suggested in this chapter have been included in the architecture, the dynamics generated internally will parallel that between the creature and its environment (compare section 4.2). It is interesting to note that the dynamical concepts introduced in chapter 4 can be used to describe both the dynamics between the organism and the environment, and the internal operation of a neural system. For example, the planning of an approach behavior would correspond to the convergence to an attractor within the artificial neural network controlling the creature. If such a mapping can be done in general, it would allow a dynamical view of cognition on all process levels. Goal-directed thought could, thus, be seen as internally performed goal-directed behavior.

9.6 Modes of Planning

Given that a creature is equipped with a planning ability of the kind described above, how and when should it be used? First we need to distinguish between two types of planning.

In the first case, planning is used to speed up learning of S–R associations by silently rehearsing stored copies of previous experiences. In this type of planning, no generalization is required in the internal environment. This is essentially the type of learning used within the DYNA framework (Sutton 1992). Stored experiences are selected at random and used to train a reinforcement leaning module. Since planning of this type is not at all goal-directed, there is no guarantee that the trained experiences will be relevant to the creature.

In the second type of planning, some form of goal-direction is involved. Such planning requires that the internal environment can be used for generalization. If the behavior that leads to the goal is already known planning is unnecessary. If it is unknown, planning is impossible. Planning, thus, requires a situation that is somewhere in between. That is, a new situation to which old knowledge can be applied.

Gulz (1991) distinguishes between two other types of planning. The first type is called *immediate* planning and is produced when the approach of a specific goal is obstructed. The planning that it generates is directed toward this specific goal. For instance, an unexpected encounter with an obstacle that is hard to overcome could generate immediate planning behavior.

The second type of planning is called *anticipatory*. Such planning does not relate to the current engagement of the creature. Instead, it is caused by simulated needs. The creature anticipates that it will become hungry tomorrow and plans its day accordingly.

There is an obvious similarity between the situation in which the orienting reaction is executed and the situations in which immediate planning could be activated. In both cases, something unexpected happens in the environment and in both cases ongoing behavior must be inhibited. This suggests that this type of planning is closely connected to emotional states. The prototypical emotion activating immediate planning is, thus, frustration caused by omission of an expected reward or by an unexpected obstacle.

Anticipatory planning, on the other hand, is not caused by any external event. It must, thus, compete with other engagements through the motivational system. This is the role of the anticipatory drive introduced in chapter 6.

There is, thus, a clear connection between emotion and immediate planning and motivation and anticipatory planning. Depending on the type of emotion or on the type of simulated motivation, we can identify different modes of planning. Different types of planning are caused by immediate punishment or frustration and by their anticipated counterparts. On the positive side, these modes are less clear, allthough one may possibly identify daydreaming with planning motivated by anticipated reward.

9.7 Conclusion

We have discussed a number of limitations of the creature that was developed in the previous chapters. It was suggested that before higher cognitive abilities can be included in the architecture, a number of problems must be solved.

First, it is necessary that the creature uses distributed representations at multiple scales and of many modalities to be able to generalize from one situation to another. This problem is addressed by adding more sensors and more sensory processing to the creature.

Second, the categorization mechanism described in chapter 5 must be extended to handle these distributed representations. The main problem here is to devise a matching criterion that generates new categories at the right times. With an incorrect matching rule, too many categories will be formed. While the mechanism described in chapter 7 can learn any expectancy, this is often done at the cost of a large number of unnecessary category nodes. For a more economical system, a more advanced mechanism for category creation is needed.

Third, the internal environment puts some non-trivial requirements on to the expectancy network. It must somehow represent both which stimulus will follow another, as well as when, and with what certainty. The need for a representation of the consequences of actions and for hierarchical chunking makes this problem even more complex. This is an area for future research.

None of the above problems appear impossible to solve, however, and in restricted cases, they all have simple solutions. Given that these problems are solved more generally, we have shown that some complex cognitive abilities will become possible. It should be clear that these systems are natural extensions of the mechanisms we have already developed.

Chapter 10

Conclusion

10.1 Introduction

This final chapter presents an overview of how the various systems described in the previous chapters can be put together into a coherent model.[1] This complete model will be viewed from three perspectives.

The next section of this chapter describes the individual components of the model and their functional relations. We will, then, go on to describe the model from an evolutionary perspective as a set of superimposed mechanisms, each of which makes the performance of the system more adaptive. Finally, we will discuss the relation between the models presented here and functional systems likely to exist in the brain.

The last sections of this chapter discuss a number of theoretical and practical limitations of the model and suggest some topics for further research.

10.2 The Components of the Model

We have tried to promote a view of cognition as a process governed by a collection of interacting systems. The main goal has been to present an overall model of how the various systems should be put together.

---

[1] To make this chapter more readable, references to chapters and sections in the rest of the book will be given without the initial words 'see section' etc.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

Behavior is seen, not as the execution of actions or responses, but as the coordinated activation and inhibition of a large number of behavior modules (3.2, 4.3, 5.2, 5.12, 6.3). Unlike responses, the behaviors controlled by behavior modules can be of arbitrary complexity (4.3, 4.4). They can range from simple responses (8.2) to complex fixed–action patterns (4.3) and goal-directed behaviors (4.2) that use both adaptation (5.7) and external stimuli for control (8.2). In some cases, the adaptation is very simple, as in the tuning of an approach behavior (5.7). In other cases, it is very complex, as, for example, when long chains of behaviors are learned (8.3). The adaptation within a behavior module can also consist of the learning of the location of a stimulus based on multiple cues (7.3).

We have divided behaviors into four classes, depending on the status of the object or situation to which it relates (4.1, 4.2, 5.2). *Appetitive* behavior is directed toward an object or a situation in the sense that it increases the probability that the goal will be found. An appetitive behavior may or may not be goal-directed. *Aversive* behavior is directed away from an object or a situation. *Neutral* behavior relates to objects that are known to be neither appetitive nor aversive. Obstacles are typically of this kind. Finally, we identified the class of *exploratory* behaviors which relate to objects or situations with unknown or uncertain valences. The valence of a stimulus always relates to a specific engagement. A stimulus that may be appetitive for one engagement may be aversive or neutral for another.

In later chapters, exploratory behavior was seen to be highly dependent on the *orienting reaction* (4.4). Efficient exploration is generated by an orienting reaction that habituates as unkown objects become aversive, appetitive or neutral (7.4).

Equipped with an appropriate behavior repertoire, the problem is to know when each behavior module should be activated or inhibited. We have described two main mechanisms for this: *learning* (5) and *motivation* (6).

Learning is the process by which an animal aquires knowledge about itself and its environment, either by trying out different behaviors in different situations (5.2, 5.9, 8.3) or by learning about contingencies in the world (5.10, 8.4).

The first type of learning was termed *procedural learning* since it established sequences of activations of behavior modules (8.3). Learning of this type is controlled by primary and secondary reward or punishment (5.2–5.6). We presented three primary events that would generate reinforcement: *presentation*, *termination* or *omission* of reward or punishment (5.2). The reinforcement generated in all these cases was seen as the result of a matching between actual and expected consequences (5.2). The difference between expected reward and the actual reward is used to generate reinforcement signals that control learning. Reinforcement is, thus, different from reward and punishment.

In *expectancy learning* predictions are made, not only of future rewards or punishments, but also of future sensory and perceptual states (5.10, 8.4). The basic learning mechanism is the same, however. The animal learns when something is unexpected. Expectancy learning is viewed as the basic mechanism behind classical conditioning, which is seen primarily as a process whereby an animal learns to predict future sensory states. Such predictions were used for different purposes in three different systems. The first was to produce responses in the traditional way (5.3). The second was to habituate the orienting reaction and the exploratory behavior (7.4). The third role was in the process we called look-ahead choice to generate recurrent predictions many states ahead (8.4).

Motivation is the second mechanism controlling the activation or inhibition of behavior modules (6). While learning adjusts the behavior of an animal to the future, motivation adjusts behavior according to its present needs. In most cases, the motivational system activates not only a single behavior module but an entire *engagement system*, consisting of a large number of interacting behavior modules and learning systems (3.2, 6.9).

The motivational system is, thus, seen as a choice mechanism that decides whichbehavior or engagement should be active at a certain time (6.1, 6.5, 6.6). This choice depends on three factors (6.2). The first is the *need* of the animal represented by its *drive signals* (6.3). The second is *external incentive*, generated by the representation of an external goal (6.4) and the third is *internal incentives*, which are generated by the representation of external cues that in some way predict the goal (6.4, 7.5). In this context *emotion*s are seen as the internal states generated by reinforcing stimuli (6.7).

Perception is considered to be a process serving at least three related functions (7.1). The first is to recognize the current situation and aid in the selection of the appropriate behavior (7.1, 7.2). The second is to guide the execution of this selected behavior (7.3). The final role of perception is to generate incentives to the motivational system which helps in the selection of an engagement (6.4, 7.5). *Categorization* plays an important role in all these functions (7.2).

We have suggested that the categorization process is driven by three different types of mismatch conditions. In the first, the best available category does not fit the input pattern sufficiently well and a new category is created (7.2, 8.3). In the second condition, the expected reward or punishment does not match the actual reward or punishment (7.1, 8.3). This implies that a finer representation of the current situation is required, which causes a new category to be formed. In the final mismatch condition, the expectations about the present situation are not matched by external stimuli (7.2). Again, a new category is required. In all three cases, a new category is only created when the mismatch is sufficiently pronounced. When it is less severe, the already existing categories or expectancies are adjusted instead.

10.3 Model Overview

Figure 10.3.1 gives an overview of the main parts of the model and their interaction. The architecture consists of a number of parallel pathways from sensors to effectors. Each of these systems can in itself control the behavior of the creature without the aid of the others, but behavior will be most effective when all systems cooperate.

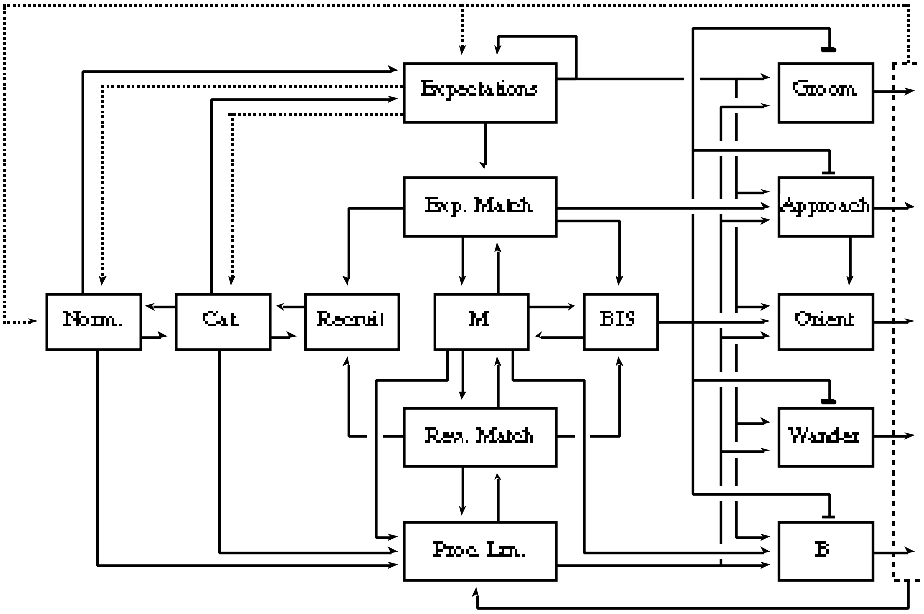


Figure 10.3.1 Overview of the proposed model and the main interaction between the various systems. Apart from the connections draw in the figure all behavior modules to the right in the figure also receive sensory input and send externala, and possibly internal, incentive signals to the motivational system. These connections are not drawn in the figure. See text for further explanation.

The Behavior Modules

The different types of behavior modules are drawn to the right in the figure. These include modules for grooming (4.4), approach (4.2), orientation (4.2, 4.4), wandering (4.4) and miscellaneous other behaviors (B, 8.3). All behavior modules are assumed to receive sensory input from the normalization system and to send external, and possibly internal, incentive (6.4) to the motivational system (M). They are also influenced by the current motivational state of the creature (6.4). These interactions are not drawn in the figure, however.

The outputs from the behavior modules pass through an arbitration section (the dashed rectangle to the right) before their output is allowed to control the motor system (4.3). The result of this arbitration is also sent to the procedural learning system (Proc. Lrn. and Rew. Match, 8.3), where it is used to drive the learning process. The behavioral selection is also influenced by the current motivational state. With lower motivation, the behavior selection is allowed to be more random due to the influence of the exploratory drive.

The Behavioral Inhibition System

The behavioral inhibition system (BIS, 4.2, 4.4, 4.5, 7.2) is responsible for the inhibition of ongoing behavior. Apart from this function, it has the effect of activating the orienting system (4.4). It also informs the motivational system that it has become active. This mechanism is activated when something unexpected happens in the environment and stops whatever the creature is doing at the moment. As a result, its sensory resources are freed and can be used to investigate the unexpected event.

The Procedural Learning System

The procedural learning system consists of one module for procedural learning (Proc. Lrn., 5.2, 5.9, 8.3) and one for matching between actual and expected rewards (Rew. Match, 5.2). The output from this system activates different behavior modules, depending on the current sensory input. It also influences the motivational system in the form of secondary internal incentive.

The procedural learning system receives inputs from four sources. The first is from the arbitration system that informs it about which behavior is currently being performed. The second is the reinforcement signals it receives from the matching system. These signals are used to strengthen the connections within the learning system. To select the appropriate behavior, the procedural learning system also receives input both from the sensory system, and from the motivational system. This allows the creature to learn to behave in a manner appropriate for its current motivational state.

The matching module compares the actual and the expected reward. If there is a severe mismatch, this module can activate the behavioral inhibition system as well as recruit new categories from the perceptual system. The matching module must also receive signals from the motivational system to be able to determine whether the perceived consequences of a behavior are appropriate or not (6.7).

The Perceptual System

The main modules of the perceptual system are the normalization system (Norm, 7.2.) which limits the dynamic range of the sensory input, the categorization system (Cat., 7.2) which generates mutually exclusive representations of its input, and the recruitment system (Recruit, 7.2) which decides when to create new categories.

The recruitment system receives input from three sources. The first comes from the categorization system. If there are no categories that match the current sensory input sufficiently well, the creation of a new category will be requested. The second input comes from matching between the actual and expected reward (Rew. Match, 5.2–5.6). When the match is too disparate, a new category is requested that will give a better representation of the current situation. The third input comes from the matching between actual and expected sensory input (Exp. Match, 5.10). This input is also used to recruit new categories.

In many cases, it is reasonable to also include the expectancy system (Expectations and Exp. Match, 5.10) and the orienting module (Orient, 4.4, 9.2) within the perceptual system, especially when backprojection from the expectancy network to the categorization and normalization systems (dashed arrows in the figure) are added to the model.

The Expectancy System

The expectancy system involves most of the modules in the system and is perhaps the clearest example of the distinction between physical modules (the boxes in the figure) and the functional systems they comprise. This system has a number of functions.

The first is in classical conditioning, where it is used to activate various behavior modules as a result of conditioning (5.3). The second function is to habituate the orienting reaction when the sensory situation becomes expected (7.4). In this process, the expectancy system interacts with the behavioral inhibition system through the expectancy matching module.

The third function is in look-ahead choice (8.4), where recurrent expectations are propagated through the expectancy system until the goal representation is activated. This process is assumed to be controlled by the expectancy matching system, which has access both to the recurrent expectations and the current motivational state. The result of this process is to activate the appropriate approach behavior (8.4). The approach system can in turn activate the orienting mechanism when the stimulus is not sufficiently straight ahead (4.2).

The fourth function of the expectancy system is to participate in the construction of perceptual categories. By continuously generating expectations about which categories should be activated when a certain stimulus is present, the expectancy matching system can recognize situations where the perceptual representation does not match expectations.

A final function of the expectancy system is to serve as an internal model used for planning and problem solving (9.4, 9.5). These processes are made possible if the dashed connections at the top of the figure are added. However, this function of the expectancy system has not yet been simulated, and many problems remains which must be solved before planning can be implemented in the model (9.4). It should be noted, however, that no specific planning module will be necessary. Planning will, thus, be a truly emergent property of the system.

The Exploratory System

Like the expectancy system, the exploratory system exploits many modules in order to accomplish its task. The orienting system directs the sensory apparatus of the creature toward novel stimuli (4.2, 4.4). The approach system lets the creature approach them (4.2). The expectancy system (5.3, 7.4) recognizes novel or missing stimuli and directs the exploration toward such situations. The motivational system (6) determines how urgent the current needs of the creature are, and this information is used to set the level of exploration that is used in selecting behavior.

In summary, the system uses a number of interacting modules to accomplish a large variety of functions.

10.4 Evolution of the Model

From an evolutionary perspective, we may view the model as a number of layers superimposed on each other. Each new layer depends on the ones already in the system to accomplish some new task. It may be instructive to imagine how this type of system could have evolved. The following is a list of a number of hypothetical evolutionary steps.

The lowest layer consists of the individual behavior modules (3.2). If these are sufficiently well–chosen, the creature is able to survive without any central control mechanism. The applicability predicates of the behavior modules decide when a particular behavior will be executed and the outputs from the modules are combined using additive composition (4.3). Adaptation within the behavior modules is already useful at this level.

The next step is to include interaction between the behavior modules as a form of arbitration (4.3). This interaction is typically some form of suppression or inhibition (4.3). It is also possible for behavior modules to generate sequences by being conected in chains (4.3).

Next, we include multi-modal interactions (6.2). We have discussed this type of interaction both within the motivational system (6.2) and within the orienting system (9.2), but it is also possible for other systems to use multi-modal input, if available. When these systems are in place, competition (3.3) can be added, both within the motivational system and within the orienting system. The competition in the motivational system selects the current motivational state (6.3) while the competition in the orienting system selects one stimulus when many are simultaneously present (4.4, 9.2).

At this point, more general learning abilities become useful. Systems for classical and instrumental conditioning can be included (5.2, 5.3). It is also possible to add an expectancy system, which interacts with the orienting system, by using a behavioral inhibition system. After this stage, one may hypothesize a number of steps in which both the expectancy and the procedural learning systems become more and more general. Starting as learning mechanisms within one behavior module, they are gradually extended to learning among larger sets of modules, until the system shown in figure 10.2.1 is reached. In this model, both the procedural and the expectancy–based learning systems can use all behavior modules instrumentally.

With added generality within the learning systems, a better perceptual system becomes useful. Again, one can image a categorization system that gradually evolves from some learning mechanism within a behavior module, for example, a place–approach system, to a more general categorization system (7.2). A related development is the inclusion of recurrent expectations for look-ahead choice (8.4).

The final step is to make the expectancy system general enough to operate as an internal environment (9.4). When this system is in place, planing and problem solving become possible (9.5, 9.6).

From an evolutionary perspective, the system starts with very specific modules and gradually develops toward a more general architecture. A fully general system is not desirable, however. A more general system requires a longer learning period before it performs as well as a specific system, and in many cases, this time is not available. This is especially the case for aversive situations (2.4).

10.5 Neural Correlates of the Model

Although it has not been the goal of this book to model any specific circuits in real brains, it may nevertheless be interesting to compare the present model with their biological counterpart. Since the model is inspired by a large range of neurophysiological findings and theories, many of the modules in figure 10.2.1 appear to have such counterparts. The systems presented here are, of course, much simpler than those in real animals, but a comparison may still be instructive. The reader should be warned not to take it too literally, though.

Starting at the input side, the normalization module is similar in some respects to the thalamus (Shepherd 1990). All inputs pass through this module, and its dynamic range is adjusted. In the real brain, olfaction is the only sensory modality that does *not* pass through thalamus, which means that a comparison is not entirely accurate, even though smell has been used as a substitute for almost all other modalities in this book.

The categorization module partly corresponds to sensory analysis in the temporal areas of the cerebral cortex, albeit in a rather impoverished version. Given the simple sensory input used in our creature, the required sensory analysis is very elementary. Most of the sensory analysis thought to take place in cortex occurs before any of the systems we have discussed in this book. We did recognize the need for the dynamic creation of categories, however. This is a function that has been ascribed to the hippocampal system (Rolls 1990). If we accept this view, the hippocampus would correspond to the recruitment module and the entorhinal cortex to the categorization module (Rolls 1990). The expectancy matching module will, then, correspond to the subiculum of the hippocampal formation (Gray 1995). This brain region has been suggested to compare the actual sensory input with the expected situation. An alternative hypothesis is that matching is performed in the CA3 area of the hippocampus (Shepherd 1990). Within Gray's theory, a mismatch within this system also activates the behavioral inhibition system (Gray 1982, 1995). A further function of the hippocampus is to handle working memory (Shepherd 1990, Squire 1992).

Procedural learning and some instances of classical conditioning are probably handled by the cerebellum (Ito 1982, Moore and Blazis 1989, Schmajuk and DiCarlo 1992). The processing of primary reward and punishment is handled by the amygdala (LeDoux 1995, Rolls 1990). The matching of actual and expected reward has been suggested to take place in the orbitofrontal cortex (Rolls 1990, 1995), which implies that this region is involved in the recognition of omitted (and possibly terminated) reward and punishment.

Some parts of the motivational system, together with behavior modules for motivational behavior, can be compared with various nuclei in the hypothalamus that have been implicated in these functions (Panksepp 1986, Schachter 1970).

The equivalent of behavior modules can be found in a number of regions in the brain. The basal ganglia may be involved in the production of approach behavior (Gray 1995). The orienting reaction is known to be controlled by the superior colliculus (Stein and Meredith 1993), although higher regions also participate (Posner and Rothbart 1992). This is also true of the behavior modules for wandering and what we called "miscellaneous behaviours" above. Large parts of these systems probably reside in the brain stem and the spinal cord (Kandel, Schwartz and Jessel 1991).

For the higher levels, comparisons with the real brain are much more difficult to make. Although various areas in the frontal cortex have been implicated in planning and problem solving, it is not at all clear what types of functions these areas contribute to the overall process (Kandel, Schwartz and Jessel 1991). Gray (1995) has suggested that the system responsible for the required predictions is the Papez loop (subicular area–mammilary bodies–anteroventral thalamus–cingulate cortex–subicular area), which in turn is auumed to receive each step in a motor program from frontal cortex via its projections to the cingulate cortex.

10.6 Theoretical Limitations

Throughout this book, the presentation has been at two different levels. On one hand, we have tried to describe a concrete artificial nervous system. This guarantees that there exists at least one functioning instance of the model we have proposed. On the other hand, we wanted the model to have more general implications for a theory of cognition. With such an approach follows the inherent danger of not making clear which claims are made about the specific artificial creature and which have larger generality.

The discussion in this section will hopefully elucidate some of these issues as we try to outline some of the theoretical limitations of the present model and give a list of problems that could, or, perhaps, should, have been addressed in the present model.

Stimulus Representation One of the largest limitations of the model is that we have assumed that there exists a small set of innately recognized stimuli. Each of these stimuli is assumed to have its own approach and orientation modules within the system. While this may be the case for a few biologically important stimuli, it cannot be true in general. The introduction of place-approach modules is a step away from this limitation, but we nevertheless assumed that each place had its own behavior module.

A more natural way to handle place-approach would be to have one single place-approach module that could approach any place. In this case, place-learning would be considered more as a perceptual process than as motor-learning. It is possible that this alternative is only a change of view, however. As we saw in chapter 7, it is not possible to draw a clear line between perception and motor control. Viewing place-learning as a perceptual process may only be a shift of this line. This view does suggest a similarity between place-learning and category-learning, however.

The matching between categories and the current sensory situation is certainly similar to the matching between a place–category and the current input. Is it perhaps possible to use the same process in both cases? If such a process can be found, it would constitute an important breakthrough, since it would essentially equate the two different types of categories.

There are also a number of limitations of the place-recognition as it was described in chapter 7. These are mainly related to the simulated olfactory modality used in our example creature. Future simulations will investigate how the model handles more realistic place categories.

Association and Generalization One important limitation of the model is that it is not able to generalize from one situation to another. This limits the power of the model considerably. It seems reasonable that generalization can be included within the model if distributed representations are used, and this is one of the most important areas for future research.

It will also be necessary to combine mechanisms for simultaneous expectations and for temporal predictions. Amit (1989) has proposed that in order to generate sequences, a neural network should include both fast symmetrical connections, corresponding to an auto-associative memory, and slow asymmetric connections that govern the transition from one state to the next. In the present model, the auto-assocaitive part would correspond to expectations established through simultaneous conditioning and by the working memory, and the slow connections would correspond to the temporal predictions made by the expectancy network.

As discussed in chapter 9, this solution is closely connected to the inclusion of more modalities and sub-modalities in the model. Many standard connectionist models handle generalization very well (Rumelhart and McClelland 1986), which implies that such abilities could easily be included in the present model. This still has to be tested, however.

Contingency vs. Contiguity One problem that the present model has in common with most accounts for conditioning is that it cannot accurately handle the difference between contingency and contiguity. Gallistel (1992) has argued that classical conditioning should be considered as an adaptive specialization for the solution of multivariate, non-stationary time-series analysis. In many cases, various statistical methods appear to model the behavior of animals better than any of the associative accounts.

It would be interesting to attempt to combine the features of these very different models. At present, the associative theories appear more realistic from a neural perspective, but some unified model will be necessary if associative theories are not to lose their credibility in the future.

This limitation relates to the somewhat limited predictions made by the present model. While simulations have been run in which an extension similar to the DYNA architecture has been used for planning, a large set of alterations are required to let the expectancy module operate as a richer internal environment (See chapter 9).

Psychological Distance The notion of psychological distance we have used is restricted to time as the only resource. This view is very simplified since it assumes that all behaviors are equally costly, and this limits the model as it stands today, in many respects. A reasonable extension would be to include a better mechanism for the calculation of psychological distance based on the actual effort needed to perform a behavior. In chapter 5, such modifications were discussed, but it is not clear how the reinforcement module can be extended to this more general case.

A related area is the discount factor that we have assumed to be constant. In a more advanced model, this factor should also represent the confidence in the prediction of future. A high discount rate would represent a high confidence in the prediction and a lower discount rate a less probable prediction. As discussed in chapter 5, the reinforcement module does represent the certainty of a prediction to some extent, but it does not influence the discount rate in the required way. Despite these limitations, the notion of psychological distance as a basis for arbitration seems very fruitful, however, and further research will show whether this is generally the case or not.

Hierarchical Learning It is well known that humans use hierarchical representations both in time and space. In the present model, this aspect of learning has been ignored. The categorization mechanism presented in chapter 7 performs a simple form of spatial chunking, but only at a single level.

To model higher cognitive processes, some form of chunking is, thus, necessary. At present, this is an area where symbolic systems perform much better than their neural network counterparts. Many questions remains to be answered before hierarchical processing becomes possible within the model.

Concurrency In the end of chapter 8, it was suggested that the expectancy system could be used as a search-heuristic for the procedural learning system. When the procedural learning system has learned a behavioral sequence sufficiently well, this model could generate routine behavior automatically without the aid of the expectancy system. It was proposed that this would allow the expectancy system to be freed for other tasks.

Although the expectancy system can certainly be disconnected when the procedural learning system has learned the task, it is not at all clear how the expectancy system could be used without the involvement of the motivational system. There appear to be two solutions. The first is to include an additional motivational system within the expectancy system, which would be used for action selection when it runs in “imagination mode”. During this period, the ordinary motivational system is used for action selection within the procedural system. The second solution is to include a local system for arbitration within the procedural learning system, which would operate without the central motivational system.

Both these solutions have a number of problems, however. How do the different arbitration mechanisms interact? How can expectancies be monitored if the expectancy network is disconnected from the sensory apparatus? Perhaps, it is not possible to address these questions within the framework for motivation presented here. However, we believe that this is possible, and will try to extend the architecture in this direction in the future.

A related problem is how the architecture can support concurrent behaviors. An obvious extension is to let the various learning systems activate a set of behavior modules rather than a single one, but this inevitably leads to a combinatorical explosion of behavioral alternatives during training. Perhaps it is possible to equip the creature initially with a set of useful motor–hierarchies (Gallistel 1980, Shepherd 1988) that support some level of concurrency, but it is not clear what these should look like.

Arousal The concept of arousal has always been a problem for the theories of learning (Gray 1975). In this book, we have chosen to ignore this important factor altogether for various reasons. The first is that arousal is intimately connected with various autonomic responses that would greatly complicate the model if they were included. The second is that arousal did not appear to be of any importance to our example creature.

To make the current model a realistic model of cognition, however, arousal cannot be ignored. The theory developed by Gray (1975, 1982) suggests a straightforward way in which to extend the model in this direction, and this will be done in the future.

Aversive Behavior A final limitation is that we have not considered aversively motivated behavior to any large extent. A more complete model must necessarily deal with fight, flight, and freeze behavior, and other emergency reactions in more detail.

10.7 Future Work

Apart from the theoretical problems described in the previous section, there are also a number of more practical limitations that will be addressed in future work. This section gives a brief overview of these future areas of research.

The analysis of the model presented in this book is far from rigorous. It would be interesting to do a mathematical analysis of the various learning systems. Under which conditions do they learn optimal behavior? When does the learning converge? How does learning-time depend on the number of behaviors and the type of stimulus-representation used?

Earlier, we developed a mathematical framework for the analysis of neural representations (Balkenius and Gärdenfors 1991, Balkenius 1992, 1994c). A long-term future goal is to extend this framework to the sequential learning mechanisms described in this book. This will hopefully make it possible to develop a model of serial compound conditioning. This is the most general case of conditioning, which involves a stimulus distributed in both time and space. Many of the possible extensions discussed in chapter 9 relate to this issue. The success of any complete theory of cognition will ultimately depend on a satisfactory account of this situation.

A more nearby goal is to adapt the proposed architecture for the control of mobile robots. This will require a number of modifications on the sensory side of the model. The most important extension is that vision will be used as the primary sensory system. This will allow us to use much more accurate place-representations than have been possible with the simple creature described in this book. A related development is to convert the neural networks proposed here into efficient algorithms. To make the presentation above as clear as possible, we have tried to use similar reinforcement modules for all tasks. In many cases, however, more efficient implementations of the individual systems are possible.

An ambitious goal would be to have a central motivational and learning system that could make use of any perceptual or behavioral system to which it is connected. The construction of a robot could then progress from a simple but operating system, to a highly complex one, simply by adding an increasing number of behavior-modules and perceptual systems. As these systems became more advanced, the more able the robot would become; and the already working system would never have to be modified. This is, of course, also the goal of the subsumption paradigm; but in such systems, the arbitration between the behavior modules is a problem for the designer of the system, not the robot itself. If this could be done automatically, the design of robots would be much easier. We think the ideas we have presented above are a step in this direction.

10.8 Final Considerations

Throughout this book, we have attempted to present many of the components that together constitute a cognitive system. Of course, we do not claim to have completely describedof any of these components. Each one is certainly much more complicated than any of the miniature models presented here.

We believe, however, that the model presented in this book can give a deeper understanding of how a large number of interacting systems can be combined. If nothing else, we hope to have contributed to the view of cognition as a process resulting from the interaction between a large number of functional systems. Unfortunately, if this view is correct, it implies that the quest for intelligence will be solved, not by the construction of one general learning algorithm, but by a lot of hard work.

Appendices

Appendix A. Simultaneous Conditioning

This appendix describes how the reinforcement module presented in chapter 5 can be used for simultaneous conditioning or for trial level explanations of conditioning. In this case, all time-delays within the reinforcement module can be ignored.

The variables shown in figure 5.2.2 are defined as follows. Let $CS = \ CS_1, CS_2, \quad , CS_n$ be a set of inputs coding the context vector (called CX in figure 5.2.2). The weight vectors that are used to calculate the activity levels of the activating node, x^+, and the inhibiting node, x^-, are called $w^+ = \ w_1^+, w_2^+, \quad , w_n^+$ and $w^- = \ w_1^-, w_2^-, \quad , w_n^-$ respectively. The reinforcement nodes are called d^+ and d^-, and their outputs are called R^+ and R^-. The inputs to the reinforcement nodes are called *Rew* and *Pun*. The following equations describe the changes in the various variables.

Let $CS(t)$ be the input to the network at time t. The activity of the activating and inhibiting nodes are calculated as,

$$x^+(t) = \sum_{i=1}^{n} CS_i(t) w_i^+(t), \qquad \text{(A1)}$$

and,

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

$$x^{-}(t) = \sum_{i=1}^{n} CS_i(t) w_i^{-}(t). \tag{A2}$$

The activity levels at the reinforcement nodes are given by,

$$d^{+}(t) = Rew(t) - x^{+}(t) + x^{-}(t) - Pun(t), \tag{A3}$$

and

$$d^{-}(t) = Pun(t) - x^{-}(t) + x^{+}(t) - Rew(t). \tag{A4}$$

The calculation of reinforcement can be compared with the well known delta rule (Widrow and Hoff 1960/1988) in which the changes to the connection weights are proportional to the difference between the target output and the actual output. In equation (A3), the desired output is represented by *Rew*(*t*) and the actual output by $x^{+}(t)$. When this difference is larger than 0, the weights will increase. When the actual output is less than the desired, the difference will yeald in increase in $d^{-}(t)$ instead as described by equation (A4). The other halves of (A3) and (A4) apply for *Pun*(*t*) and $x^{-}(t)$. The calculated differences are used to calculate the output from the reinforcement nodes as,

$$R^{+}(t) = [d^{+}(t)]^{+}, \tag{A5}$$

and

$$R^{-}(t) = [d^{-}(t)]^{+}, \tag{A6}$$

where $[x]^{+} = \max(0, x)$. The weights change according to the equations,

$$w_i^{+}(t+1) = w_i^{+}(t) + \gamma^{+} R^{+}(t) CS_i(t), \tag{A7}$$

and,

$$w_i^{-}(t+1) = w_i^{-}(t) + \gamma^{-} R^{-}(t) CS_i(t). \tag{A8}$$

In these equations, γ^{+} and γ^{-} are constants which describe the learning rate of the activating and inhibiting sides of the reinforcement module, respectively. The relation to the delta rule can be seen by setting the target output, *T*(*t*), to,

$$T(t) = Rew(t) - Pun(t), \tag{A9}$$

and the actual output *A*(*t*) to,

$$A(t) = x^{+}(t) - x^{-}(t). \tag{A10}$$

and substituting a single weight, $w_i(t)$, for the difference $w_i^{+}(t) - w_i^{-}(t)$. The delta rule can now be written as,

$$\Delta w_i(t) = CS_i(t)(T(t) - A(t)). \tag{A11}$$

The difference $T(t)$–$A(t)$, or the delta, is thus equivalent to,

$$\frac{1}{2}(d^+(t) - d^-(t)) \tag{A12}$$

in equation (A3) and (A4). In this formulation, the reinforcement module is, thus, essentially an implementation of the Rescorla–Wagner model (Rescorla and Wagner 1972) that incorporates the idea of different systems for reward and punishment (Mowrer 1960, Gray 1975). As described in chapter 5, this makes it possible to distinguish between unknown stimuli, and stimuli with net reward of 0, which is important for the habituation of the orienting reaction (See chapter 7).

Appendix B. Temporal and Secondary Conditioning

To model temporal predictions, the equations in appendix A must be changed in the following way. Replace equation (A3) and (A4) with,

$$d^+(t) = Rew(t) – x^+(t-1) + x^-(t-1) – Pun(t), \tag{B1}$$

and

$$d^-(t) = Pun(t) – x^-(t-1) + x^+(t-1) - Rew(t). \tag{B2}$$

The learning equations can now be reformulated to only change the weights after CS–onset. To accomplish this, let $\Delta CS_i(t)$ represent the onset of CS_i, that is,

$$\Delta CS_i(t) = [CS_i(t) - CS_i(t-1)]^+. \tag{B3}$$

Equations (A7) and (A8) can now be replaced by,

$$w_i^+(t+1) = w_i^+(t) + \gamma^+ R^+(t) \Delta CS_i(t), \tag{B4}$$

and,

$$w_i^-(t+1) = w_i^-(t) + \gamma^- R^-(t) \Delta CS_i(t). \tag{B5}$$

These equations will result in conditioning when the the onset of a CS precedes a reward or punishment by exactly on time step. Appendix C describes how the reinforcement system can be changed to allow conditioning also at other interstimulus intervals.

For second order conditioning, it is necessary that all inputs to the reinforcement module can drive the learning process. This is accomplished by adding the current output of the activating and inhibiting nodes to the activity of the reinforcement nodes (See also figure 5.6.2):

$$d^+(t) = Rew(t) - x^+(t-1) + \delta x^+(t) + x^-(t-1) - Pun(t) - \delta x^-(t), \tag{B6}$$

and

$$d^-(t) = Pun(t) - x^-(t-1) + \delta x^-(t) + x^+(t-1) - Rew(t) - \delta x^+(t). \tag{B7}$$

As described in chapter 5, the constant δ represents the discount rate for future events or rewards.

Appendix C. Interstimulus Interval Effects

In the main text, we have not discussed the role of interstimulus interval effects. The learning equations from appendix B can easily be changed to handle such effects, however. To do this, we need to introduce the concept of an *eligibility trace* (Klopf, Morgan and Weaver 1993). Such a trace represents that a weight is eligible for change during an interval of time after the onset of its corresponding CS. The eligibility is assumed to decay with time, which accounts for the different levels of conditioning obtained at different interstimulus intervals. Let $E_i(t)$ be the eligibility trace of CS_i at time t. This quantity is calculated as,

$$E_i(t) = \sum_{k=1}^{m} c_k \Delta CS_i(t-k), \tag{C1}$$

where c_k are constants that describe the strength of conditioning with interstimulus interval of k time units. Klopf, Morgan and Weaver (1993) suggested that, to describe the conditioning found in various experiments, these constants should be set as shown in table C1 (See also section 5.11).

Constant	Values
c_1	5.0
c_2	2.5
c_3	1.25
c_4	0.625
c_5	0.3125

Table C1 The values for the constants c_i as described by Klopf, Morgan and Weaver (1993).

The following equations should be substituted for equations (B4) and (B5) to model interstimulus interval effects.

$$w_i^+(t+1) = w_i^+(t) + \gamma^+ R^+(t) E_i(t), \tag{C2}$$

and,

$$w_i^-(t+1) = w_i^-(t) + \gamma^- R^-(t) E_i(t). \tag{C3}$$

Allthough these changes to the learning equations are fairly simple, they do introduce a number of complex questions in relation to serial compound conditioning. Since a discussion of these questions would led us too far from the topic of this book, we have chosen not to include this extension in the main text.

Appendix D. Sequential Learning

To implement sequential learning, the reinforcement module described in appendix A and B must be extended with a set of facilitated connections as shown in figure 5.9.2 and 8.3.1. Let w_{ij}^+ be the connection from stimulus representation ΔCS_i, facilitated by the behavior selection node b_j, to the activating side of the reinforcement module, and let w_{ij}^- be the corresponding connection to the inhibitory side. The weight w_{ij} represents the connection from the stimulus representation for CS_i to the behavior selection node b_j.

The weights w_{ij}^+ and w_{ij}^- change according to equations (B4) and (B5) above when b_j is active. That is, the weights only change when they are facilitated. In other cases, they are kept constant. The weights w_{ij} change according to,

$$w_{ij}(t+1) = w_{ij}(t) + \gamma(x^+(t) - x^+(t))\Delta CS_i b_j. \tag{D1}$$

In the simulations presented in chapter 8, the winner of the behavioral competition among the b_j nodes was set to 1 directly, and the activity of all other nodes were set to 0. The winner was selected from a Bolzmann-distribution as described by equation 4.3.2.

Appendix E. Expectancies and Look-Ahead Choice

Expectancy learning uses one reinforcement module, as defined in appendix A and B, for each stimulus representation. The expectancy network is thus entirely specified by the equations in those appendices.

To use the expectancy network for look-ahead choice, as shown in figure 8.4.1, the following procedure is used. The calculation of the best goal-predictor initially clears the state of the choice mechanism *u*, and then performs the following calculation for each perceivable stimulus, CS_i, around the creature: At t=0, v_i is set to 1 and the reinforcement module for CS_i receives an input of 1. For a limited number of time-steps, 10 in the simulations, activity is propagated through the expectancy network. If during this time, the node *p* is activated, the state, *u*, of the choice mechanism is updated as follows:

$$u_i(t+1) = v_i(t) p_i(t) \tag{E1}$$

In other cases, it is held constant. When the goal-prediction of each stimulus has been computed, the node a is activated to choose the best goal-predictor. In the simulations, the node a_i with the largest value on u_i was activated directly, and the output of all other nodes where set to 0. The output of a_i is then used to select the appropriate behavior.

Appendix F. Category Recruitment

This appendix describes the competitive recruitment learning presented in section 7.2 in detail. The variables were also defined in that section.

All constants can be choosen freely in the intervals described here. The slope of output function $f(x)$ is controlled by $m > 1$ and $0 < \gamma < 1$. The resting potential of the nodes r and d is $\rho > 0$. The update rate of the integrating nodes is ε, where $0 < \varepsilon << 1$. The learning rate is δ, where $0 < \varepsilon << \delta$. The weights w_{ij} are initially set at random in the range $[0, a]$, where $0 \le a < 1$. The weights v_i are set at random in the range $[b, c]$, where $0 < b < c < 1$. All other variables are initially set to 0. Three output functions are needed for the network:

$$f(x) = \frac{(x-\theta)^m}{\gamma + (x-\theta)^m}, \tag{F1}$$

$g(x) = \max(0, x)$, and $h(x)$ is the threshold function. Given an input pattern $s = s_0, s_1, \ldots, s_n$ the signals in the network are calculated according to the following equations:

$$x_i(t+1) = \left[x_i(t) + \varepsilon\left(\sum_{j=1}^{n} s_j w_{ji}(t) - g(x_0(t))\right)\right]^+ \tag{F2}$$

$$x_0(t+1) = \sum_{i=1}^{n} f(x_i(t)) \tag{F3}$$

$$r(t+1) = \rho - g(x_0(t)) \tag{F4}$$

$$y(t+1) = \rho - h(r(t)) \tag{F5}$$

$$y_i(t+1) = [y_i(t) + \varepsilon(h(r(t))) - g(y_0(t))]^+ \tag{F6}$$

$$y_0(t+1) = \sum_{i=1}^{n} f(y_i(t)) \tag{F7}$$

For each time step, the weights in the network are updated as follows:

$$w_{ij}(t+1) = w_{ij}(t) - \delta f(y_j(t))[s_i - w_{ij}(t)] \tag{F8}$$

$$v_i(t+1) = v_i(t) - \delta f(x_i(t))h(d(t))v_i(t) \tag{F9}$$

These equations will automatically create new categories when needed, or activate old ones when this is appropriate. The resting potential ρ of the node r controls when new categories must be created. The larger this value is, the more categories will be formed. An external input to r can be used to dynamically set this level, that is, to create new categores on demand.

Appendix G. Iterative Normalization

This appendix describes the iterative normalization presented in section 5.2. Let $s = \; s_1, s_2, \quad , s_n$ be the input signals to the nodes $x_1 \ldots x_n$, and, let b be the activity of the bias node. It is further assumed that,

$$f(x) = g(x) = h(x) = [x]^+. \tag{G1}$$

The outputs, $f(x_i)$, from the nodes x_i are multiplied with the bias and summed at x_0:

$$x_0(t+1) = \sum_{i=1}^{n} h(b(t))f(x_i(t)) \tag{G2}$$

The output from node x_0, $g(x_0)$ either increases or decreases the activity of the node b depending on whether it is larger or smaller than 1:

$$b(t+1) = [b(t) + \varepsilon(1 - g(x_0(t)))]^+. \tag{G3}$$

To use the euclidean norm, we have to sum the square of the outputs instead. This can be accomplished by adding a self-facilitating connection to the output of each x_i. See figure (N1). We also need to replace the output function of x_0 with,

$$g(x) = \sqrt{x}. \tag{G4}$$

Given that the constant ε is sufficiently small, these equations will iteratively normalize an input vector using either the Euclidaen or city-block norm.

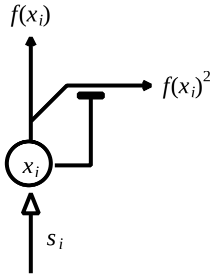


Figure G1 Calculation of the square of a signal using self-faciliation.

Appendix H. Place Approach

In place approach, as described in section 7.3, one reinforcement module is used for each smell receptor of the creature. By using the reinforcement module for simultaneous conditioning, as described in appendix A, and using a large learning rate, the creature can quickly establish a representation of a place. To let the module learn the smell configuration at a specific place, a learning signal, *L*, is used to facilitate the reinforcement connections within the module (See figure 7.3.1), that is, the learning equation (A7) and (A8) are replaced by,

$$w^+_{ijk}(t+1) = w^+_{ijk}(t) + \gamma^+ L_j R^+(t) s_{ik}(t), \tag{H1}$$

and,

$$w^-_{ijk}(t+1) = w^-_{ijk}(t) + \gamma^- L_j R^-(t) s_{ik}(t), \tag{H2}$$

where *i* is the index of the smell receptor, *j* the place, and *k* the side of the creature (left or right). The behavioral output of the module for side *k* and place *j* is calculated as,

$$m^k = \sum_{i=1}^{n} (d^+_{ijk}(t) + d^-_{ijk}(t)), \tag{H3}$$

where *n* is the number of sensory receptors.

From a technical standpoint, this is simply a very complicated mechanism for computing the distances between the prototype vector w_j for place *j*, and the current sensory inputs at the left and the right, s^L and s^R. An alternative way to construct a place–approach module is, thus, to set the prototype of place *j* to the average of s^L and s^R when L_j is active, and then compute the output of place unit *j* as,

$$m^k = \sum_{i=1}^{n} \left| w_{ijk} - s_i^k \right|. \tag{H4}$$

Alternatively, the Euclidean distance could have been used instead. In a creature with visual abilities, it would be more advantageous to use visual input rather than smell. In this case, the smells would be replaced by the visual angles of a number of distal landmarks (Compare Zipser 1985, Schmajuk and Blair 1983).

Appendix I. Exploration

This appendix describes the details of the exploration network shown in figure 7.4.1. The expectancy part of the network uses the reinforcement module described in appendix A. Instead of using the actual intensity of each smell input, CS_i is set to 1, if smell i is present, and to 0 otherwise. To choose which stimulus to approach, each orientation and approach module weigh together three factors. The first is the the ‘unexpectancy’ factor which is $CS_i T(d_i^+ - \alpha)$, where T is the threshold function. The second is the ‘misplace’ factor, $\gamma T(d_i^- - \beta)$. Here, α and β control when a stimulus is sufficiently unexpected or misplaced to activate the orienting system. The constant γ determines the salience of a missing stimulus. The last factor is the ‘novelty’ of a stimulus that models habituation within the orientation module. The habituation of module i is represented by the value h_i, which changes according to the following equation when module i is active:

$$h_i(t+1) = h_i(t) + \delta CS_i(t)(1 - h_i(t)) \tag{I1}$$

The ‘novelty’ factor is computed as $(1 - h_i)(\varepsilon T(CS_i) + CS_i)$. Here, the constant δ describes how fast the approach behavior habituates, and ε controls the salience of a novel stimulus. Note that equation (I1) describes habituation that is not reversible. To be more realistic, the habituation should slowly decrease to zero again.

The three factors described here are added together in each orienting module, and is used as a strength output in the behavioral competition between these modules. In the simulations described in chapter 7, the orienting modules also compete with a wandering module that wins the competition when the strength outputs of the orienting modules are to low.

Appendix J. The Simulator

All the models presented in this book have been implemented as computer simulations. Since many of the models involves dynamic situations, viewing the behavior of the creatures ’live’, rather than as static images, makes it much easier to understand how they behave. The dynamics of the interaction between a creature and its environment simply cannot be given justice in the format offered by a book. To remedy this situation, the simulator, “BERRY”, which runs on Macintosh™ computers, is available to anyone interested. It can be downloaded from the world-wide web or obtained directly from the author. To obtain the simulator though world-wide web, open,

http://lucs.fil.lu.se/Simulators/BERRY-III/

and follow the instructions on–line. The simulator can also be obtained directly from the author, who can be contacted either by e-mail to,

christian.balkenius@fil.lu.se

(place the word ‘BERRY’ in the subject line) or by ordinary mail to,

Christian Balkenius
Lund University Cognitive Science
Kungshuset/Lundagård
S–222 22 LUND, Sweden

References

Abbott, E. A., (1884/1991), *Flatland*, Princeton, NJ: Princeton Univerity Press.

Ackley, D. H., Hinton, G. E. & Sejnowski, T. J., (1985), “A learning algorithm for Bolzmann machines”, *Cognitive Science*, **9**, 147–169.

Adams, D. B., (1979), “Brain mechanisms for offense, defense, and submission”, *The Behavioral and Brain Sciences*, **2**, 201–243.

Agre, P. E. & Chapman, D., (1987), “Pengi: an implementation of a theory of activity”. *Proceedings of the Sixth National Conference on Artificial Intelligence*, **1**, 268–272, Morgan Kaufman.

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.

Albus, J. S., (1991), “Outline for a theory of intelligence”, *IEEE Transactions on Systems, Man and Cybernetics*, **21**, 473–509.

Amari, S.–I., (1982), “Competitive and cooperative aspects in dynamics of neural excitation and self-organization”. In S.–I. Amari and M. A. Arbib (eds.), *Competition and Cooperation in Neural Nets*, 1–28, Berlin: Springer–Verlag.

Amari, S.–I. & Arbib, M. A., (1977), “Competition and cooperation in neural nets”. In J. Metzler (ed.), *Systems Neuroscience*, 119–165, Academic Press.

Amit, D. J., (1989), *Modelling Brain Function*, Cambridge: Cambridge University Press.

Christian Balkenius, 1995, Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37, 1995.

Andersen, R. A. & Zipser, D., (1990), “A network model for learned spatial representation in the posterior parietal cortex”. In J. L. McGaugh, N. M. Weinberger and G. Lynch (eds.), *Brain Organization and Memory: Cells, Systems, and Circuits*, 271–284, Oxford: Oxford University Press.

Arbib, M. A. & Cobas, A., (1991), “Schemas for prey-catching in frog and toad”. In J.–A. Meyer and S. W. Wilson (eds.), *From Animals to Animats*, 540–549, Cambridge, MA: MIT Press.

Arbib, M. A., Conklin, E. J. & Hill, J., (1987), *From Schema Theory to Language*, Oxford University Press.

Arbib, M. A. & House, D. H., (1987), “Depth and detours: an essay on visually guided behavior”. In M. A. Arbib and A. R. Hanson (eds.), *Vision, Brain and Cooperative Computation*, 129–163, Cambridge, MA: MIT Press.

Arbib, M. A. & Lee, H. B., (1993), “Anural visuomotor coordination for detour behaviour: from retina to motor schemas”. In J.–A. Mayer, H. L. Roitblat and S. W. Wilson (ed.), *From Animals to Animats II*, Cambridge, MA: MIT Press.

Arkin, R. A., (1990), “Integrating behavioral, perceptual and world knowledge in reactive navigation”. In P. Maes (ed.), *Designing Autonomous Agents*, 105–122, Cambridge, MA: MIT Press.

Arkin, R. C., (1987), “Motor schema based navigation for a mobile robot: an approach to programming by behavior”. (ed.), *Proceedings of the 1987 IEEE International Conference on Robotics and Automation*, 264–271.

Atkeson, C. G. & Rinkensmeyer, D. J., (1990), “Using associative content-addressable memories to control robots”. In P. H. Winston and S. A. Shellard (eds.), *Artificial Intelligence at MIT: Expanding Frontiers*, Cambridge, MA: MIT Press.

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

Balkenius, C. & Gärdenfors, P., (1991), “Non-monotonic inferences in neural networks”. In J. A. Allen, R. Fikes and E. Sandewall (eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, 32-39, San Mateo, CA: Morgan Kaufman.

Balkenius, C., (1992), “Neural mechanisms for self-organization of emergent schemata, dynamical schema processing, and semantic constraint satisfaction”, *Lund University Cognitive Studies*, **14**.

Balkenius, C., (1993), “The roots of motivation”. In J.–A. Mayer, H. L. Roitblat and S. W. Wilson (eds.), *From Animals to Animats II*, Cambridge, MA: MIT Press.

Balkenius, C., (1994a), “All you never wanted to know about the BERRY III environment”, *LUCS Minor*, **2**.

Balkenius, C., (1994b), "Natural intelligence for autonomous agents – extended version", *Lund University Cognitive Studies*, **29**. (Also printed in B. Svensson (ed.), *Proceedings of the International Workshop on Mechatronical Computer Systems for Perception and Action*, 1993, 181–196.)

Balkenius, C., (1994c), "Some properties of neural representations". In L. F. Niklasson and M. B. Bodén (eds.), *Connectionism in a Broad Perspective*, 79–88, New York: Ellis Horwood.

Balkenius, C., (1995), "Multi-modal sensing for robot control". In L. F. Niklasson and M. B. Bodén (eds.), *Current Trends in Connectionism*, 203–216, Hillsdale, NJ: Lawrence Erlbaum.

Barto, A. G., Anderson, C. W. & Sutton, R. S., (1982), "Synthesis of nonlinear control surfaces by a layered associative search network", *Biological Cybernetics*, **43**, 175–185.

Barto, A. G. & Sutton, R. S., (1981), "Landmark learning: an illustration of associative search", *Biological Cybernetics*, **42**, 1–8.

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.

Beer, D. B., (1990), *Intelligence as Adaptive Behavior*, San Diego: Academic Press.

Beer, R. A., Chiel, H. J., Quinn, R. D. & Larsson, P., (1992), "A distributed neural network architecture for hexapod robot locomotion", *Neural Computation*, **4**, 356–365.

Beer, R. D. & Gallagher, J. C., (1992), "Evolving dynamical neural netwoks for adaptive behavior", *Adaptive Behavior*, **1**, 92–122.

Bell, W. J., (1991), *Searching Behaviour*, London: Chapman and Hall.

Bolles, R. C., (1967), *Theory of Motivation*, New York: Harper and Row.

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.

Bolles, R. C., (1978), "The role of stimulus learning in defensive behaviour". In S. H. Hulse, H. Fowler and W. K. Honig (eds.), *Cognitive Processes in Animal Behavior*, 89–108, Hillsadale, NJ: Lawrence Erlbaum Associates.

Bolles, R. C., (1984), "Species-typical response predispositions". In P. Marler and H. S. Terrace (eds.), *The Biology of Learning*, 435–446, Berlin: Springer–Verlag.

Braitenberg, V., (1984), *Vehicles: Experiments in Synthetic Psychology*, Cambridge, MA: MIT Press.

Brooks, R. A., (1986), “Achieving artificial intelligence through building robots”, 899, MIT A. I. Memo.

Brooks, R. A., (1990), “Elephants don't play chess”. In P. Maes (ed.), *Designing Autonomous Agents*, 3–15, Cambridge, MA: MIT Press.

Brooks, R. A., (1991a), “Intelligence without reason”. *Proceedings of IJCAI–91*, 569–595.

Brooks, R. A., (1991b), “New approaches to robotics”, *Science*, **253**, 1227–1232.

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

Brown, M. W., (1990), “Why does the brain have a hippocampus?”. In M. Gabriel and J. Moore (eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, 233–282, Cambridge, MA: MIT Press.

Buck, R., (1988), *Human Motivation and Emotion*, New York: John Wiley & Sons.

Cannon, W. B., (1932), *The Wisdom of the Body*, New York: Norton.

Carpenter, G. A. & Grossberg, S., (1986), “Neural dynamics of category learning and recognition: attention, memory consolidation and amnesia”. In S. Grossberg (ed.), *The Adaptive Brain*, 238–308, Amsterdam: North–Holland.

Carpenter, G. & Grossberg, S., (1987), “ART2: Self-organization of stable category recognition codes for analog input patterns”, *Applied Optics*, 26, 4919-4930.

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

Cecconi, F. & Parisi, D., (1993), “Neuronal networks with motivational units”. In J.–A. Mayer, H. L. Roitblat and S. W. Wilson (eds.), *From Animals to Animats II*, 346–355, Cambridge, MA: MIT Press.

Charniak, E. & McDermott, D., (1985), *Introduction to Artificial Intelligence*, Reading, MA: Addison–Wesley.

Churchland, P. M., (1984), *Matter and Consciouness: A Contemporary Introduction to the Philosophy of Mind*, Cambridge, MA: MIT Press.

Clark, F. C., (1958), “The effect of deprivation and frequency of reinforcement on variable–interval responding”, *Journal of the Experimental Analysis of Behavior*, **1**, 211–227.

Cohen, M. A. & Grossberg, S., (1983), “Absolute stability of global pattern formation and parallel memory storage by competitive neural networks”, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC–13**, 815–826.

Cohen, M. A. & Grossberg, S., (1987), “Masking fields: a massively parallel neural architecture for learning, recognition, and predicting multiple groupings of patterned data”, *Applied Optics*, 26, 1866-1891.

Collier, G. H., (1983), “Life in a closed economy: the ecology of learning and motivation”. In M. D. Zeiler and P. Harzem (ed.), *Advances in Analysis of Behavior*, 223–274, New York: John Wiley.

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.

Davis, J. L. & Eichenbaum, H., (1991), *Olfaction*, Cambridge, MA: MIT Press.

Desmond, J. E., (1990), “Temporally adaptive responses in neural models: the stimulus trace”. In M. Gabriel and J. Moore (eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, 421–456, Cambridge, MA: MIT Press.

Deutsch, J. A., (1960), *The Structural Basis of Behavior*, Chicago: University of Chicago Press.

DeYoe, E. A. & Van Essen, D. C., (1988), “Concurrent processing streams in monkey visual cortex”, *Trends in Neurosciences*, **11**, 219–226.

Dickinson, A., (1980), *Contemporary Animal Learning Theory*, Cambridge: Cambridge University Press.

Dunham, P. J., (1977), “The nature of reinforcing stimuli”. In W. K. Honig and J. E. R. Staddon (ed.), *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.

Ekeberg, Ö., (1992), *Computer Simulation Techniques in the Study of Neural Systems*, Stockholm: Royal Institute of Technology, Dept. of Numerical Analysis and Computing Science.

Eldén, L. & Wittmeyer-Koch, L., (1987), *Numerisk Analys – En Introduktion*, Lund: Studentlitteratur.

Ellen, P. & Thinus-Blanc, C., (1987), *Cognitive Processes and Spatial Orientation in Animal and Man*, Dordrecht: Nijhoff.

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.

Flavell, J. H., (1985), *Cognitive Development*, New Jersey: Prentice-Hall.

Gallistel, C. R., (1980), *The Organization of Action: A New Synthesis*, Hilldsdale, NJ: Erlbaum.

Gallistel, C. R., (1990), *The Organization of Learning*, Cambridge, MA: MIT Press.

Gallistel, C. R., (1992), “Classical conditioning as an adaptive specialization: a computational model”, In *The Psychology of Learning and Motivation: Advances in Research and Theory*, Medin, D. L., (ed.),, 35–67, New York: Academic Press.

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

Garcia, J., McGowan, B. K. & Green, K. F., (1972), "Biological constraints on conditioning". In A. H. Black and W. F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory*, 3–27, New York: Appleton–Century–Crofts.

Gärdenfors, P., (1992), "Medvetandets evolution". *Blotta Tanken,* ch 5, Nya Doxa.

Gärdenfors, P., (1995), "Cued and detached representations in animal cognition", *to appear in Behavioural Processes*.

Gärdenfors, P. & Balkenius, C., (1993), "Varför finns det inga riktiga robotar?", *Framtider*, **12**, 11–14.

Gibbon, J., Baldock, M. D., Locurto, C., Gold, L. & Terrace, H. S., (1977), "Trial and intertrial durations in autoshaping", *Journal of Experimental Psychology Animal Behavioral Processes*, **3**, 264–284.

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.

Goldberg, M. E., Eggers, H. M. & Fouras, P., (1991), "The ocular motor system", In E. R. Kandel, J. H. Schwartz and T. M. Jessel (eds.), *Principles of Neural Science*, 660–678, New York: Elsevier.

Gray, J. A., (1975), *Elements of a Two-Process Theory of Learning*, London: Academic Press.

Gray, J. A., (1982), *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*, Oxford: Oxford University Press.

Gray, J. A., (1995), "A model of the limbic system and the basal ganglia: applications to anxiety and schizophrenia", In Gazzaniga, M. S. (ed.),, *The Cognitive Neurosciences*, 1165–1176, Cambridge, MA: MIT Press.

Grelle, M. J. & James, J. H., (1981), "Conditioned inhibition of fear: evidence for a competing response mechanism", *Learning and Motivation*, **12**, 300–320.

Grossberg, S., (1973), "Contour enhancement, short term memory, and constancies in reverbating neural networks", *Studies in Applied Mathematics*, **LII**, 213–257.

Grossberg, S., (1976), "Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors", *Biological Cybernetics*, **23**, 121–134.

Grossberg, S., (1986), "The adaptive self-organization of serial order in behavior: speech, language, and motor control". In E. C. Schwab and H. C. Nusbaum (eds.), *Pattern Recognition by Humans and Machines*, Academic Press.

Grossberg, S., (1989), "Nonlinear neural networks: principles, mechanisms, and architectures", *Neural Networks*, **1**, 17–66.

Grossman, S. P., (1979), "The biology of motivation", *Annual Review of Psychology*, **30**, 209–242.

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

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

Harnad, S., (1987), “The symbol grounding problem”, *Physica D*, **42** , 335-346.

Hinton, G. E. & Andersson, J. A., (1981), *Parallel Models of Associative Memory*, Hillsdale, N.J.: Erlbaum.

Hinton, G. E., McClelland, J. L. & Rumelhart, D. E., (1986), “Distributed representations”. In D. E. Rumelhart and J. L. McClelland (eds.), *Parallel distributed processing*, Cambridge, MA: MIT Press.

Hirose, S., (1993), *Biolgically Inspired Robots*, Oxford: Oxford University Press.

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

Hopfield, J. J., (1984), “Neurons with graded response have collective computational properties like those of two-state neurons”, *Proceedings of the National Academy of Sciences*, **81** ,

Horswill, I. D., (1992), “Characterizing adaptation by constraints”. In F. J. Varela and P. Bourgine (eds.), *Towards a Practice of Autonomous Systems*, 58–63, Cambridge, MA: MIT Press.

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

Horswill, I. D. & Stein, L. A., (1995), “Life after planning and reaction”, *manuscript*, MIT A. I. Lab.

Hull, C. L., (1932), “The goal-gradient hypothesis and maze learning”, *Psychological Review*, **39** , 25–43.

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

Hull, C. L., (1938), “The goal-gradient hypothesis applied to some “field-force” problems in the behaviors of young children”, *Psychological Review*, **45** , 271–299.

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.

Johnston, A. J. & Dark, V. J., (1986), “Selective attention”, *Annual Review of Psychology*, **37** , 43–75.

Kaelbling, L. P., (1993), *Learning in Embedded Systems*, Cambridge, MA: MIT Press.

Kamin, L. J., (1968), “Attention-like processes in classical conditioning”. In M. R. Jones (ed.), *Miami Symposium on the Prediction of Behavior: Aversive Stimulation*, 9–31, Miami: University of Miami Press.

Kamp, Y. & Hasler, M., (1990), *Recursive Neural Networks for Associative Memory*, Chichester: John Wiley & Sons.

Kandel, E. R., (1991), “Perception of motion, depth, and form”. In E. R. Kandel, J. H. Schwartz and T. M. Jessel (eds.), *Principles of Neural Science*, 440–466, New York: Elsevier.

Kandel, E. R. & Jessel, T. M., (1991), “Touch”. In E. R. Kandel, J. H. Schwartz and T. M. Jessel (eds.), *Principles of Neural Science*, 367–384, New York: Elsevier.

Kandel, E. R., Schwartz, J. H. & Jessel, T. M., (1991), *Principles of Neural Science*, New York: Elsevier.

Kehoe, E. J., (1990), “Classical conditioning: fundamental issues for adaptive network models”. In M. Gabriel and J. Moore (eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, 390–420, Cambridge, MA: MIT Press.

Khatib, O., (1985), “Real-time obstacle avoidance for manipulators and mobile robots”. *Proceedings of the 1985 IEEE International Conference on Robotics and Automation*, 500–505.

Kiss, G., (1991), “Autonomous agents, AI and chaos theory”. In J.–A. Meyer and S. W. Wilson (ed.), *From Animals to Animats*, 518–524, Cambridge, MA: MIT Press.

Klopf, A. H., (1988), “A neuronal model of classical conditioning”, *Psychobiology*, **16** , 85–125.

Klopf, A. H. & Morgan, J. S., (1990), “The role of time in natural intelligence: implications of classical and instrumental conditioning for neuronal and neural-network modeling”. In M. Gabriel and J. Moore (eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, 463–495, Cambridge, MA: MIT Press.

Klopf, A. H., Morgan, J. S. & Weaver, S. E., (1993), “A hierarchical network of control systems that learn: modelling nervous system function during classical and instrumental conditioning”, *Adaptive Behavior*, **1**, 263–319.

Kohonen, T., (1984), *Self-Organization and Associative Memory.*, Berlin: Springer Verlag.

Kohonen, T., Oja, E. & Lehtiö, P., (1989), “Storage and processing of information in distributed associative memory systems”. In G. E. Hinton and J. A. Anderson (eds.), *Parallel Models of Associative Memory*, 125–167, Hillsdale, NJ: Lawrence Erlbaum.

Konorski, J., (1967), *Integrative Activity of the Brain*, Chicago: University of Chicago Press.

Kraft, L. G. & Campagna, D. P., (1990), “A summary comparison of CMAC neural network and traditional adaptive control systems”. In W. T. Miller, R. S.

Sutton and P. J. Werbos (eds.), *Neural Networks for Control*, 143–169, Cambridge, MA: MIT Press.

Lansner, A. & Ekeberg, Ö., (1989), "A one-layer artificial neural network with a Bayesian learning rule", *International Journal of Neural Systems*, **1**, 77–87.

LeDoux, J. E., (1995), "In search of an emotional system in the brain: leaping from fear to emotion and consciousness". In M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, 1049–1061, Cambridge, MA: MIT Press.

Lee, T., (1970), "Perceived distance as a function of direction in the city", *Envrionmental Behavour*, **2**, 40–51.

Lewin, K., (1935), *A Dynamic Theory of Personality: Selected Papers*, New York: McGraw–Hill.

Lewin, K., (1936), *Principles of Topological Psychology*, New York: McGraw–Hill.

Lieberman, D. A., (1990), *Learning: Behavior and Cognition*, Belmont, CA: Wadsworth Publishing Company.

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

Livingstone, M. & Hubel, D., (1988), "Segregation of form, color, movement, and depth: Anatomy, physiology, and perception", *Science*, 740–749.

Lorenz, K., (1950), "The comparative method in studying innate behavior patterns", Society for Experimental Biology, Symposium no. 4, *Physiological Mechanisms of Animal Behavior*, 221–268, New York: Academic Press.

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

Lubow, R. E. & Moore, A. U., (1959), "Latent inhibition: the effect of nonreinforced preexposure to the conditioned stimulus", *Journal of Comparative Physiological Psychology*, **52** , 415–419.

Lynch, G. & Granger, R., (1991), "Serial steps in memory processing: possible clues from studies of plasticity in the olfactory–hippocampal circuit". In J. L. Davis and H. Eichenbaum (eds.), *Olfaction*, 141–165, Cambridge, MA: MIT Press.

Lyons, D. M., Hendriks, A. J. & Mehta, S., (1991), "Achieving robustness by casting planning as adaptation of a reactive system". *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*,

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

Maes, P., (1990), *Designing Autonomous Agents*, Cambridge, MA: MIT Press.

Maes, P., (1991), "A bottom-up mechanism for behavior selection in an artificial creature". In J.–A. Meyer and S. W. Wilson (eds.), *From Animals to Animats*, 238–246, 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.

Martin, J. H., (1991), "Coding and processing of sensory information". In E. R. Kandel, J. H. Schwartz and T. M. Jessel (eds.), *Principles of Neural Science*, 329–340, New York: Elsevier.

Mason, C. & Kandel, E. R., (1991), "Central Visual Pathways". In E. R. Kandel, J. H. Schwartz and T. M. Jessel (eds.), *Principles of Neural Science*, New York: Elsevier.

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.

McFarland, D., (1993), *Animal Behaviour*, Harlow: Longman.

McFarland, D. & Bösser, T., (1993), *Intelligent Behavior in Animals and Robots*, Cambridge, MA: MIT Press.

McFarland, D. & Sibly, R., (1972), " 'Unitary drives' revisited", *Animal Behavior*, **20**, 548–563.

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

Miller, N. E., (1944), "Experimental studies of conflict". In J. M. Hunt (ed.), *Personality and the Behavior Disorders*, New York: The Ronald Press Co.

Minsky, M., (1987), *The Society of Mind*, London: Heineman.

Minsky, M. & Papert, S., (1988), *Perceptrons – expanded edition*, Cambridge, MA: MIT Press.

Mishkin, M. & Petri, H. L., (1986), "Memories and habit: some implications for the analysis of learning and retention". In L. R. Squire and N. Butters (eds.), *Neuropsychology of Memory*, 287–296, New York: Guildford Press.

Mishkin, M., Ungerleider, L. G. & Macko, K. A., (1983), "Object vision and spatial vision: two cortical pathways", *Trends in Neuroscience*, **6**, 414–417.

Moore, J. W. & Blazis, D. E. J., (1989), "Cerebellar implementation of a computational model of classical conditioning". In P. Strata (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.

Mowrer, O. H., (1960/1973), *Learning Theory and Behavior*, New York: Wiley.

Murray, E. A., (1990), “Representational memory in nonhuman primates”. In R. P. Kesner and D. S. Olton (eds.), *Neurobiology of Comparative Cognition*, Hillsdale, NJ: Lawrence Erlbaum.

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

Newell, A. & Simon, H. A., (1963), “GPS, a program that simulates human thought”. In E. A. Feigenbaum and J. Feldman (eds.), *Computers and Thought*, New York: McGraw–Hill.

Nilsson, N. J., (1994), “Teleo-reactive programs for agent control”, *Journal of Artificial Intelligence Research*, **1**, 139–158.

Norman, D. A., (1990), “Four (more) issues for Cognitive Science”, *Report 9001*, Department of Cognitive Science, San Diego, La Jolla, CA.

O’Keefe, J. & Nadel, L., (1978), *The Hippocampus as a Cognitive Map*, Oxford: Clarendon Press.

Olton, D. S., Branch, M. & Best, P., (1978), “Spatial correlates of hippocampal unit activity”, *Experimental Neurobiology*, 387–409.

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

Ortony, A., Clore, G. L. & Collins, A., (1988), *The Cognitive Structure of Emotions*, Cambridge: Cambridge University Press.

Panksepp, J., (1986), “The neurochemistry of behavior”, *Annual Review of Psychology*, **37** , 77–107.

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.

Peitgen, H. O., Jürgens, H. & Saupe, D., (1992), *Chaos and Fractals: New Frontiers of Science*, Berlin: Springer-Verlag.

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

Pitz, G. F., & Sachs, N. J., (1984), “Judgement and decisions: Theory and application”, *Annual Review of Psychology*, 35, 139-163.

Plutchik, R., (1991), *The Emotions*, Lanham: University Press of America.

Powers, W. T., (1973), *Behavior: The Control of Perception*, Chicago: Aldine Publishing Company.

Posner, M. I. & Rothbart, M. K. (1992), “Attention regulation: from mechanism to culture”, In P. Bertelson, P. Elen and G. d’Ydewalle, G., (eds.), *Current Advances in Psychological Science: Ongoing Research*, Hove: Lawrence Erlbaum.

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 generalization: reinforcement and punishment”. In R. Glazer (ed.), *On the Nature of Reinforcement*, New York: Academic Press.

Prescott, T. J. & Mayhew, E. W., (1993), “Building long-range cognitive maps using local landmarks”. In J.–A. Mayer, H. L. Roitblat and S. W. Wilson (ed.), *From Animals to Animats II*, 233–242, Cambridge, MA: MIT Press.

Raskin, D. C., (1972), “Orienting and defensive reflexes and conditioning”. In A. H. Black and W. F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory*, 269–289, New York: Appleton–Century–Crofts.

Rescorla, R. A., (1980), *Pavlovian Second–Order Conditioning: Studies in Associative Learning*, New York: Erlbaum/Wiley.

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.

Roitblat, H. L., (1994), “Mechanism and process in animal behavior: models of animals, animals as models”. In D. Cliff, P. Husband, J.–A. Meyer and S. W. Wilson (eds.), *From Animals to Animats III*, Cambridge, MA: MIT Press.

Rolls, E. T., (1986), “A theory of emotion, and its application to understanding the neural basis of emotion”. In Y. Oomura (ed.), *Emotions: Neural and Chemical Control*, 325–344, Tokyo: Japan Scientific Societies Press.

Rolls, E. T., (1990), “Functions of neuronal networks in the hippocampus and of backprojections in the cerebral cortex in memory”. In J. L. McGaugh, N. M. Weinberger and G. Lynch (eds.), *Brain Organization and Memory: Cells, Systems, and Circuits*, 184–210, Oxford: Oxford University Press.

Rolls, E. T., (1995), “A theory of emotion and consciousness, and its application to understanding the neural basis of emotion”. In M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, 1091–1106, Cambridge, MA: MIT Press.

Rosch, E. H., (1973), “Developmental changes in memory and the acquisition of language”. In T. Moore (ed.), *Cognitive Development and the Acquisition of Language*, 111–144, New York: Academic Press.

Rosen, R., (1985), *Anticipatory Systems – Philosophical, Mathematical and Methodological Foundations*, Elmsford, N.Y.: Pergamon Press.

Rosenbleuth, A., Wiener, N. & Bigelow, J., (1943), “Behavior, purpose, and teleology”, *Philosophy of Science*, **10** , 18–24.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J., (1986), “Learning internal representations by error propagation”. In D. E. Rumelhart and J. L. McClelland (eds.), *Parallel Distributed Processing*, 318–362, Cambridge MA: MIT Press.

Rumelhart, D. E. & McClelland, J. L., (1981), “Interactive processing through spreading activation”. In A. M. Lesgold and C. A. Perfetti (eds.), *Interactive Processes in Reading*, Hillsdale, NJ: Erlbaum.

Rumelhart D. E., and McClelland, J. L., (1986), *Parallel Distributed Processing*, Cambridge MA: MIT Press.

Rumelhart, D. E. & Zipser, D., (1986), “Feature discovery by competative learning”. In D. E. Rumelhart and J. L. McClelland (eds.), *Parallel Distributed Processing*, 151–193, Cambridge MA: MIT Press.

Runeson, S., (1989), “A note on the utility of ecological incomplete invariants”, *Newsletter of the International Society for Ecological Psychology*, **4**, 6–9.

Sahota, M. K., (1994), “Action selection for robots in dynamic environments through inter-behaviour bidding”. In D. Cliff, P. Husbands, J.–A. Meyer and S. W. Wilson (eds.), *From Animals to Animats III*, 138–142, Cambridge, MA: MIT Press.

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

Schachter, S., (1964), “The interaction of cognitive and physiological determinants of emotional state”. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, New York: Academic Press.

Schachter, S., (1970), “Some extraordinary facts about obese humans and rats”, *American Psychologist*, **26**, 129–144.

Schmajuk, N. A., (1990), “Role of the hippocampus in temporal and spatial navigation: an adaptive neural network”, *Behavioral Brain Research*, **39**, 205–229.

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. & DiCarlo, J. J., (1992), “Stimulis configuration, classical conditioning, and hippocampal function”, *Psychological Review*, **99**, 268–305.

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

Schnepf, U., (1991), “Robot ethology: a proposal for the research into intelligent autonomous systems.”. In J.–A. Meyer and S. W. Wilson (eds.), *From Animals to Animats*, 465–474, Cambridge, MA: MIT Press.

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

Shepard, (1984), “Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming”, *Psychological Review*, **91**, 417–447.

Shepherd, G. M., (1988), *Neurobiology*, Oxford: Oxford University Press.

Shepherd, G. M., (1990), *The Synaptic Organization of the Brain*, New York: Oxford Univeristy Press.

Shepherd, G. M., (1991), "Computational structure of the olfactory system". In J. L. Davis and H. Eichenbaum (eds.), *Olfaction*, 3–41, Cambridge, MA: MIT Press.

Shimamura, A. P., (1990), "Forms of memory: issues and directions". In J. L. McGaugh, N. M. Weinberger and G. Lynch (eds.), *Brain Organization and Memory: Cells, Systems, and Circuits*, 159–173, Oxford: Oxford University Press.

Simmons, J. A., (1989), "A view of the world through the bat's ear: the formation of acoustic images in echolocation", *Cognition*, **33** , 155–199.

Sjölander, S., (1993), "Fulguration, Dezentrierung und inneres Probehandeln. Evolutionäre Durchbrüche in der Evolution der Sprache und Kommunikation, im Sinne von Lorenz, Piaget und Freud.", *Wiener Beiträge zur Wissenschafftstheorie*, **5**.

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.

Skinner, B. F., (1974), *About Behaviorism*, New York: Vintage Book.

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

Smith, M. C., Coleman, S. R. & Gormezano, I., (1969), "Classical conditioning of the rabbit's nictitating membrane response at backward, simultaneous, and forward CS–US interval", *Journal of Comparative and Physiological Psychology*, **69** , 226–31.

Solomon, R. L., (1980), "The opponent process theory of acquired motivation", *American Psychologist*, **35** , 691–712.

Sperling, G., (1960), "The information available in brief visual presentations", *Psychological Monographs*, **74** ,

Squire, L. R., (1992), "Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans", *Psychological Review*, **99** , 195–231.

Stein, B. E. & Meredith, M. A., (1993), *The Merging of the Senses*, Cambridge, MA: MIT Press.

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.

Teitelbaum, P., (1961), “Disturbances of feeding and drinking behavior after hypothalamic lesions”. In M. R. Jones (ed.), *Nebraska Symposium on Motivation*, Lincoln: University of Nebraska Press.

Tenenberg, J., Karlsson, J. & Whitehead, S., (1993), “Learning via task decomposition”. In J.–A. Mayer, H. L. Roitblat and S. W. Wilson (eds.), *From Animals to Animats II*, Cambridge, MA: MIT Press.

Thelen, E. & Smith, L. B., (1994), *A Dynamical Systems Approach to the Development of Cognition and Action*, Cambridge, MA: MIT Press.

Thrun, S. B., (1992), “Efficient exploration in reinforcement learning”, CMU–CS–92–102, Carnegie Mellon University.

Timberlake, W., (1983), “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.

Tinbergen, N., (1951/1989), *The Study of Instinct*, Oxford: Oxford University Press.

Tolman, E. C., (1932), *Purposive Behavior 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.

Tortora, G. J. & Anagnostakos, N. P., (1990), *Principles of Anatomy and Physiology*, New York: Harper Collins.

Touretzky, D. S. & Redish, A. D., (1995), “Landmark arrays and the hippocampal cognitive map”. In L. F. Niklasson and M. B. Bodén (eds.), *Current Trends in Connectionism*, 1–13, Hillsdale, NJ: Lawrence Erlbaum.

Trehub, A., (1991), *The Cognitive Brain*, Cambridge, MA: MIT Press.

Tyrell, T., (1993), *Computational Mechanisms for Action Selection*, Unpublished doctoral thesis, Centre for Cognitive Science, University of Edinburgh, Scotland.

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

Werner, G. M., (1994), “Using second order neural connections for motivation of behavioral choices”. In D. Cliff, P. Husbands, J.–A. Meyer and S. W. Wilson (eds.), *From Animals to Animats III*, 154–161, Cambridge, MA: MIT Press.

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.

Williams, R. J., (1986), “The logic of activation functions”. In D. E. Rumelhart and J. L. McClelland (eds.), *Parallel Distributed Processing*, 423–443, Cambridge, MA: MIT Press.

Witkin, A., (1983), “Scale-space filtering”, *8th International Joint Conference on Artificial Intelligence*, 1019-1022.

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